Statistical Analysis Plan

Title:	A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Parallel Group, Multi-Center Study to Investigate the Safety and Efficacy of APD334 in Patients with Moderately to Severely Active Ulcerative Colitis
Document Number:	APD334-003
Product Name:	APD334
Formulation:	Capsules
Indication:	Ulcerative Colitis
Sponsor:	Arena Pharmaceuticals, Inc. 6154 Nancy Ridge Drive San Diego, CA 92121
Authors:	[redacted] PPD Global Ltd.
Approval:	[redacted] PPD [redacted] Arena Pharmaceuticals, Inc
Date:	27 Feb 2018

NCT Number: NCT02447302
This NCT number has been applied to the document for purposes of posting on clinicaltrials.gov

TABLE OF CONTENTS

List of In-Text Tables

[redacted]

List of Appendices

[redacted]

[redacted]

[redacted]

LIST OF ABBREVIATIONS

µg	microgram
ANCOVA	analysis of covariance
BLQ	below limit of quantification
CBC	complete blood count (test)
CI	confidence interval
CRF	case report form
CMH	Cochran-Mantel-Haenszel
CRO	contract research organization
DMC	Data Monitoring Committee
DSMB	Data Safety Monitoring Board
ECG	electrocardiogram
eCRF	electronic case report form
FCS	fully conditional specification
[redacted]	
ITT	intent-to-treat
LOCF	Last Observation Carried Forward
LSM	least square mean
MAR	missing at random
MCS	Mayo Clinic Score
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MITT	modified intent-to-treat
[redacted]	
MNAR	missing not at random
PGA	Physician's Global Assessment
[redacted]	
PMS	Partial Mayo Score
POC	Proof of concept
PRO	patient reported outcome
q.d.	once daily
QT_c	QT interval corrected for HR
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SOC	System Organ Class
TEAE	treatment-emergent adverse event
TMS	Total Mayo Score
WHODRUG	World Health Organization Drug Dictionary

1 INTRODUCTION

1.1 Objective of the Statistical Analysis Plan

This statistical analysis plan (SAP) provides the statistical rationale and methods that will be applied to data gathered in clinical trial Protocol No.APD334-003[1] in order to assess the safety and efficacy of APD334 in patients with moderately to severely active ulcerative colitis. Section 9 of this SAP discusses the changes in the planned analysis from the protocol. This SAP is finalized prior to database lock and data analysis start. Major changes in the analysis that are made after database lock will be documented in the Clinical Study Report with the rationales and details.

Unless otherwise noted, all statistical testing is one-sided and it is appropriate for this proof-of-concept phase 2 study to assess treatment effects of APD334 in patients with moderately to severely active ulcerative colitis.

1.2 Description of the Study and Objectives/Hypotheses

1.2.1 Summary of Study Design

APD334-003 is a phase 2, proof-of-concept and dose ranging clinical study designed to test the safety and efficacy of APD334 in patients with moderately to severely active ulcerative colitis and who have demonstrated, over the previous 5 year period, an inadequate response to, loss of response to, or intolerance of at least one of the following agents: oral 5-aminosalicylates, corticosteroids, immunosuppressives, TNFα antagonists, or integrin antagoinists. Eligible patients will be randomized into a double-blind, placebo-controlled study to receive once daily (q.d.) doses of APD334 (1 mg or 2 mg) or matching placebo in a 1:1:1 ratio for 12 weeks and will be stratified by presence or absence of current oral corticosteroid usage and previous exposure to TNFα antagonists. Patients who are receiving corticosteroids at baseline must remain on the same dose for the duration of the study. No more than 50% of patients will have been previously exposed to TNFα antagonists. The trial will include adult men and women, ages 18-80 years, who have moderately to severely active ulcerative colitis (defined as a 3 component Mayo score (including the endoscopic, stool frequency, and rectal bleeding) of 4 to 9 with an endoscopic subscore of ≥ 2 and a rectal bleeding score of ≥ 1) [redacted]

The total study duration is approximately 18 weeks; 4 weeks for screening procedures, followed by 12 weeks of dosing, and a possible follow-up visit at week 14. Patients will capture daily stool frequency and rectal bleeding using an electronic device throughout the study and will return to the study at regular intervals for safety, [redacted], and efficacy assessments. A 34-week extension study is offered to patients who wish to continue treatment (APD334-005). Patients who choose not to participate in the extension study or who have discontinued prematurely from the study will have a 2-week follow-up visit after the last clinical visit. The schedule of procedures and visits for the study is provided in [redacted].

1.2.2 Efficacy Questionnaire: Mayo Clinic Score (MCS)

1.2.2.1 Complete Mayo Clinic Score (Total Mayo Score, TMS)

The complete Mayo Clinic Score is an instrument designed to measure disease activity of ulcerative colitis and consists of 4 subscores: stool frequency, rectal bleeding, findings of flexible proctosigmoidoscopy, and physician global assessment and total score ranging from 0 to 12 (see Appendix 3). Each component is scored individually on an integer scale of 0 to 3 (0=normal, 1=mild, 2=moderate, 3=severe), with higher scores indicating greater disease activity.

At Screening Visit:

The MCS will be evaluated during screening using patient diary entries within the 10 days prior to randomization and flexible proctosigmoidoscopy results within 10 days prior to randomization. The subscores for stool frequency and rectal bleeding are derived from the patient diaries. The scores from the 3 most recent days prior to the actual day of the study visit will be averaged and rounded to the nearest integer. The rounding will be applied to each subscore prior to the creation of the total score. Note that the day prior, day of and day after proctosigmoidoscopy cannot be used for patient diary entry because of the required bowel prep for the procedure. Patients who have less than 3 days of diary data during screening are not eligible for randomization. Examples of the subscore entries for eligibility and subscore derivation are provided in Table 1.

Table 1. Examples of eDiary Subscore Entries for Study Eligibility and Corresponding Subscore Derivation

Example	Diary Day											Valid Days for Calculation of Subscore	Average Subscore	Final Subscore
	-10	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**1**			
Diary #1	1	2	3	X	P	X	2	**1**	**3**	**1**	E	-1, -2, -3	1.66	2
Diary #2	2	2	2	**3**	X	P	X	M	**1**	**3**	E	-1, -2, -7	2.33	2
Diary #3	3	2	3	**3**	**3**	X	P	X	**3**	E		-2, -6, -7	3.00	3
Diary #4	**3**	**0**	**1**	X	P& E	X						-8, -9, -10	1.33	1
Diary #5	1	2	X	P&E	X							Missing	N/A	Missing
Diary #6	1	2	3	1	3	3	**3**	**2**	**3**	X	P&E	-2, -3, -4	2.67	3

Abbreviations: E = eligibility; M = missing; P = proctosigmoidoscopy; X = non-scoring day before and after proctosigmoidoscopy

The physician's global assessment (PGA) acknowledges the three other criteria findings of the MCS, the patient's daily record of abdominal discomfort and general sense of well-being, and other observations, such as physical findings and the patient's performance.

At Other Study Visits:

For each study visit during the study (excluding the screening/baseline visit which uses 10 prior days), stool frequency and rectal bleeding subscores will be derived from electronic patient diaries completed over the 7 days prior to a study visit. Note that the day prior, day of and day

after proctosigmoidoscopy will not be used for patient diary entry because of the required bowel prep for the procedure. These subscores will be calculated using the following rules:

1. The scores from the 3 most recent days prior to the actual day of the study visit will be averaged and rounded to the nearest integer.
2. If patient diary entries from 3 days are not available, the scores from the 2 most recent entries will be averaged and rounded to the nearest integer.
3. If less than 2 days of diary data are available, the subscore will be considered missing.

The rounding will be applied to each subscore prior to the creation of the total score. Table 2 provide examples of the subscore entries and subscore derivation.

Table 2. Examples of eDiary Subscore Entries and Corresponding Subscore Derivation

Example	Diary Day[1]							Valid Days for Calculation of Subscore	Average Subscore	Final Subscore
	-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**			
Diary #1	1	2	3	1	0	0	3	-1, -2, -3	1.00	1
Diary #2	X	P	X	2	1	3	1	-1, -2, -3	1.66	2
Diary #3	1	2	3	X	P	X	2	-1, -5, -6,	2.33	2
Diary #4	2	M	M	1	M	M	M	-4, -7	1.5	2
Diary #5	M	M	3	M	M	M	M	Missing	N/A	Missing

[1] Days are named relative to the Day of Study Visit; Abbreviations: M = missing; P = proctosigmoidoscopy; X = non-scoring day before and after proctosigmoidoscopy

The MCS will also be evaluated at Week 12 using the Week 12 proctosigmoidoscopy and stool frequency and rectal bleeding scores completed by the patients seven days prior to the visit.

The proctosigmoidoscopy subscore is evaluated by the investigator and a blinded central reader. For analysis purposes, only the central reader results be will used and the derived Mayo Clinic Scores calculated by case report form (CRF) Health eDiary vendor will be analysed.

1.2.2.2 3-Component Mayo Clinic Score (Partial Mayo Score, PMS)

There are two 3-component Mayo Clinic Scores consisting of 3 of the 4 subscores found in the MCS.

1. PMS#1: the primary endpoint, consists of subscores for stool frequency, rectal bleeding and findings of flexible proctosigmoidoscopy.
2. [redacted]

Total score for the 3-component MCS range: 0 to 9, each component ranging from 0 to 3 (0=normal, 1=mild, 2=moderate, 3=severe). Both 3-component Mayo scores will be evaluated at baseline and Week 12.

1.2.2.3 2-Component Mayo Clinic Score

The 2-component Mayo Clinic Score consists of 2 of the 4 subscores found in the MCS (i.e. rectal bleeding and findings on endoscopy). Total score for the 2-component MCS range: 0 to 6, each component ranging from 0 to 3 (0=normal, 1=mild, 2=moderate, 3=severe). The 2-component Mayo score will be evaluated at baseline and Week 12.

[redacted]

1.2.3 Efficacy Questionnaire: [redacted]

[redacted]

1.2.4 Objectives

Primary Objective: The primary objective of this proof-of-concept study is to determine the effect of treatment with APD334 in improving PMS#1, 3-component Mayo Clinic Score (score ranging from 0 to 9, including stool frequency, rectal bleeding and findings on endoscopy) at Week 12

Secondary Objectives:

- To determine the effect of treatment with APD334 on endoscopic improvement at 12 weeks
- To determine the effect of APD334 treatment on 2-component Mayo Clinic Score (rectal bleeding and findings on endoscopy) at 12 weeks
- To determine the effect of APD334 treatment Total Mayo Clinic Score at 12 weeks

Medical Rationale for Choice of Secondary Endpoints

- Endoscopic improvement of mucosal inflammation, as assessed by proctosigmoidoscopy, is a secondary endpoint for this study, and is considered a lead indicator of endoscopic mucosal healing in a proof-of-concept clinical trial. This endpoint measure is selected because colorectal mucosal inflammation is a defining, diagnostic feature of ulcerative colitis that provides direct, visual evidence of the inflammation characteristic of the disease[2], and a key goal in treating patients with ulcerative colitis is to control inflammation[3,4]. Mucosal healing is considered an important treatment goal in both clinical trials and clinical practice because it is associated with improved outcomes, including sustained clinical remission, steroid-free remission, decreased rates of surgery, and fewer hospitalizations[4]. Endoscopic assessments of mucosal inflammation are encouraged by the Food and Drug Administration (FDA) according to the interim approaches to efficacy assessments described in the current Draft Guidance for Industry[5]. The European Medicines Agency (EMA) requires demonstration of treatment effects on the inflammatory process in ulcerative colitis by endoscopic assessments, as described in the Agency's Draft Guideline[6].

- A two-component Mayo Clinic Score (selecting the "Rectal bleeding" and "Findings on flexible sigmoidoscopy" components, and omitting the "Stool frequency" and "Physician's global assessment" components as originally described[7]) is a secondary endpoint for this study. The rectal bleeding component is selected because of the clarity and ease of patients' reports of this finding on a contemporaneous basis. The findings on flexible sigmoidoscopy component is selected based on the rationale provided in the preceding paragraph. The stool frequency component is omitted because the scoring method requires patients to assess contemporaneous stool frequency in comparison to that during a previous "normal" health state, either before symptoms of ulcerative colitis began or during periods of remitted and quiescent disease. As such, the "normal" stool frequency is subject to patient recall bias. Further, the difficulties in patients' capabilities in accurately reporting numbers of stools necessitates specific instruction, as described in the FDA Guidance[5], and which unfortunately may not be heeded or consistently followed by investigators or understood by patients. The physician's global assessment component is omitted because, per the FDA Guidance[5], clinician-reported outcome measures cannot directly assess symptoms that are known only to the patient, and FDA encourages sponsors to develop patient-reported outcome (PRO) instruments as an ideal primary efficacy assessment tool for the treatment of ulcerative colitis[5].

- Total Mayo Score (4-component and ranging 0 to 12), will reflect a broad examination of potential clinical improvement across four domains (stool frequency, rectal bleeding, findings on flexible proctosigmoidoscopy and physician's global assessment).

Safety Objective:

To determine the safety profile and tolerability of APD334 induction treatment

Exploratory Objectives:

- To determine the effect of treatment with APD334 in inducing clinical remission at 12 weeks
- To determine the effect of treatment with APD334 in inducing clinical response at 12 weeks
- To determine the effect of treatment with APD334 on a combination of clinical remission and clinical response reflected by a composite endpoint at 12 weeks

1.2.5 Hypotheses

1.2.5.1 Primary Hypothesis

In patients with moderately to severely active ulcerative colitis, treatment with APD334 compared with placebo will provide greater improvement in 3-component Mayo Clinic Score (consisting of subscores for stool frequency, rectal bleeding and findings of endoscopy) at 12 weeks.

1.2.5.2 Secondary Hypotheses

- In patients with moderately to severely active ulcerative colitis, treatment with APD334 compared with placebo will provide more patients that achieve endoscopic improvement at 12 weeks.

- In patients with moderately to severely active ulcerative colitis, treatment with APD334 compared with placebo will provide greater improvement in 2-component Mayo Clinic Score (consisting of rectal bleeding and findings on endoscopy) at 12 weeks.

- In patients with moderately to severely active ulcerative colitis, treatment with APD334 compared with placebo will provide greater improvement in Total Mayo Clinic Score (consisting of stool frequency, rectal bleeding, findings of flexible proctosigmoidoscopy, and physician global assessment) at 12 weeks.

2 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

All baseline patient characteristics of demographic data (age, height, and weight), ulcerative colitis history, social history, medical history (abnormalities only), physical examination (abnormalities only), and concomitant medications at study entry will be listed for all patients. Demographic and baseline characteristics of the safety population will be summarized by treatment group and for the overall population. Baseline for the demographic variables will be the last pre-randomization value collected at the screening visit. Continuous variables will be summarized using number of values (n), mean, standard deviation (SD), median, minimum, and maximum. Frequencies and percentages will be reported for all categorical data. No formal statistical testing comparing treatment groups will be performed.

The following variables will be summarized by treatment group:

1. Continuous baseline demographic variables: age (years).
2. Categorical baseline demographic variables: sex (female or male), race (American Indian or Alaska Native, Asian, Black or African-American, Native Hawaiian or Other Pacific Islander, White or Caucasian, or Other) and ethnicity (Hispanic or Latino, Not Hispanic or Latino).
3. Ulcerative colitis history which include:
   - Duration of Ulcerative Colitis (in years)
   - Disease localization (Proctosigmoiditis/Left sided colitis, Pancolitis/ Extensive colitis)
   - Presence or absence of current oral corticosteroid usage at baseline
   - Inadequate responders/Lost response and/or Intolerant to corticosteroid (including duration of treatment)
   - Previous exposure to TNFα antagonists (including duration of treatment)
   - Inadequate responders/Lost response and/or Intolerant to TNFα antagonists
   - Previous exposure to immunosuppressives (including duration of treatment)
   - Inadequate responders/Lost response and/or Intolerant to immunosuppressives
   - Previous use of integrin antagonist (including duration of treatment) and response
   - Previous use of oral 5-aminosalicylates (including duration of treatment) and response
   - Baseline Disease Activity (Mayo Clinic Scores)

3 PATIENT DISPOSITION

The number of patients enrolled in the study by country and treatment group will be tabulated. Tables showing study participant accounting will be provided. Tables will indicate number of patients who were randomized into the study, the number of patients who completed treatment and/or completed the study, and the number of patients who discontinued treatment and/or discontinued study prematurely (early termination) for any of the following reasons:

- Adverse event(s)
- Patient lost to follow up
- Patient withdrawal of consent
- Investigator decision
- Sponsor decision
- Death
- Other

Protocol deviations will be listed and tabulated by subtype.

4 EFFICACY ANALYSES

4.1 Efficacy Endpoints

A table is provided in Appendix 2 summarizing the efficacy variables and their analysis populations.

Primary Endpoint

Improvement of PMS#1, 3-component Mayo Clinic Score (score ranging from 0 to 9, including stool frequency, rectal bleeding and findings on endoscopy) at Week 12.

Secondary Endpoints

Secondary efficacy variables include:

- Proportion of patients who achieve endoscopic improvement at Week 12

  Endoscopic improvement is defined as Mayo endoscopic subscore (using findings of flexible proctosigmoidoscopy) of ≤ 1 point.

- Improvement in 2-component Mayo score (score ranging from 0 to 6, including rectal bleeding and findings on endoscopy) at Week 12

- Improvement in Total Mayo Clinic Score (score ranging from 0 to 12, including stool frequency, rectal bleeding, findings on endoscopy and physician's global assessment) at Week 12.

Exploratory Endpoints

The exploratory efficacy endpoints are listed below:

- The trichotomous composite endpoint of clinical remission and clinical response (score ranging 0 to 2: score 2 for achieving both clinical remission and clinical response; 1 for achieving only clinical response, and 0 for achieving neither) at Week 12.
- The proportion of patients achieving clinical remission at Week 12

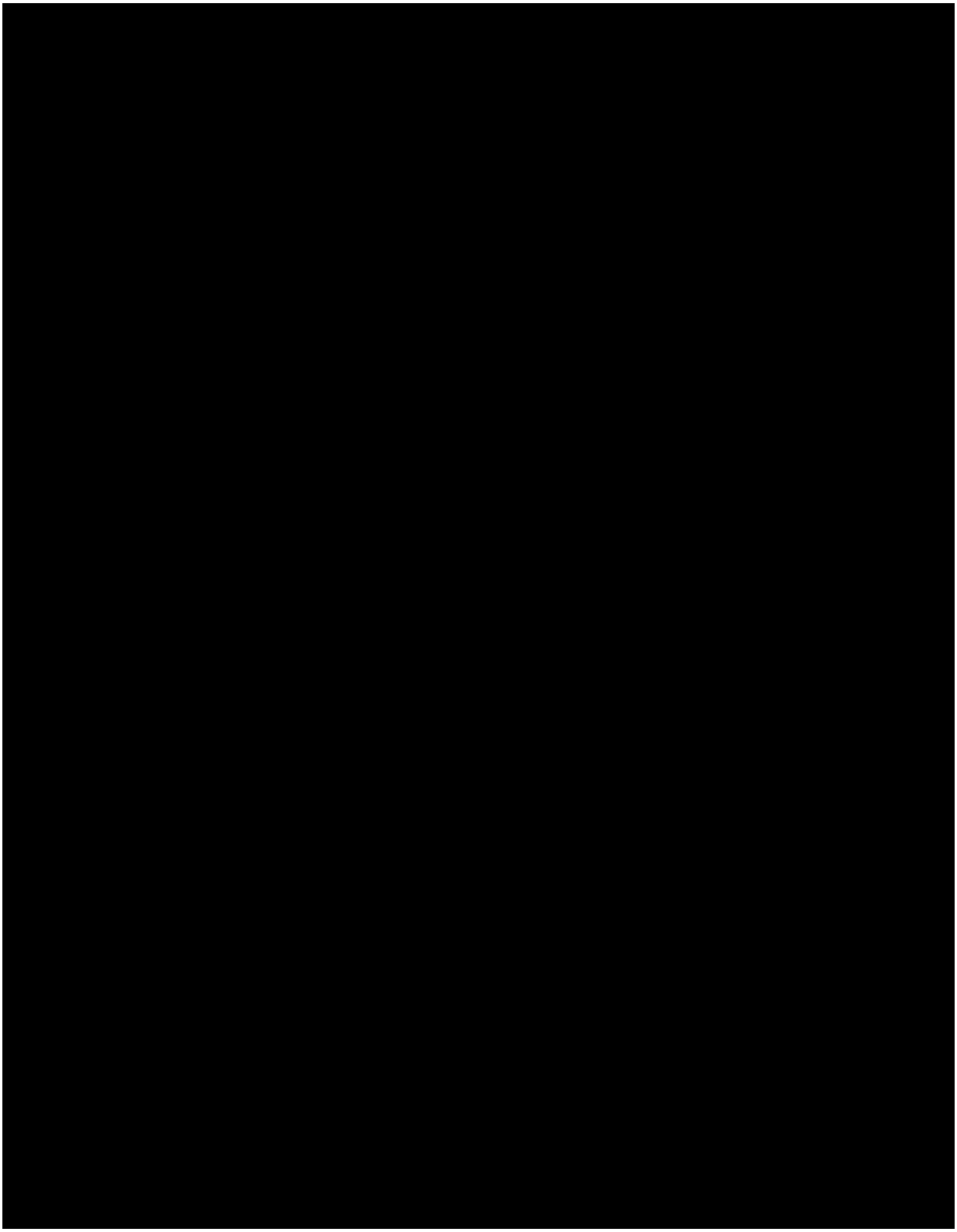
  A patient has achieved clinical remission if he/she has:

  - an endoscopy score using flexible proctosigmoidoscopy of 0 or 1 (excluding friability) and
  - a rectal bleeding score of 0 or 1 and
  - a stool frequency score of 0 or 1 with a decrease of ≥ 1 point from baseline

- Proportion of patients who achieve clinical response at Week 12

  A patient has achieved clinical response if he/she meets criteria of clinical remission defined above, or meet criteria of clinical response defined in the protocol. Clinical

response is defined as a decrease in 3-component Mayo Clinic score of ≥ 2 points and a decrease of ≥ 30% with either a decrease of rectal bleeding of ≥ 1 or rectal bleeding score of 0 or 1



4.2 Analysis Population

All analyses on efficacy variables will use the Modified Intent-to-Treat (MITT) population as primary. A completer's population will be used as a secondary analysis population for the primary and secondary endpoints. The Intent-to-Treat (ITT) population will be used as sensitivity analysis population for proportion-based efficacy endpoints.

Intent-to-Treat Population (ITT):

This population consists of all randomized patients, who received at least 1 dose of study medication. Under this approach, patients are counted in the treatment group to which they were randomized, regardless of the treatment received during the course of the trial. In this population, missing data handling for efficacy analyses will be specified in Section 7.4.

Modified Intent-to-Treat Population (MITT):

This population consists of all randomized patients, who received at least 1 dose of study medication, have a baseline measurement, and have a post-randomization measurement. Under this approach, patients are counted in the treatment group to which they were randomized, regardless of the treatment received during the course of the trial. Note that MITT population can vary with endpoints since some patients may have the needed data for inclusion in the MITT population for some endpoints but not for others.

Completers Population (CP):

This population consists of all patients in the MITT population who completed the study. No missing data will be imputed for this analysis. Any substantial differences between conclusions based on the ITT /MITT population and the completers' population will be investigated.

Safety Population (SP):

The Safety Population will include all randomized patients who received study medication.

4.3 Approaches to Efficacy Analysis

The efficacy endpoints at Week 12 will be analysed in the ITT population using the multiple imputation (MI) method to handle missing data. See Section 7.4.1, Section 9 and Appendix 5 for details.

For all analyses, two models will be fitted. One model will be fitted for the individual treatment comparisons (APD334 2mg vs Placebo, APD334 1mg vs Placebo, APD334 2mg vs APD334 1mg) with the treatment term containing the three treatment groups, namely APD334 2mg, APD334 1mg and Placebo, and a second model will be fitted for the pooled treatment comparison (APD334 2mg & 1mg vs Placebo) with the treatment term containing two treatment groups, pooled APD334 2mg & 1mg and Placebo.

4.3.1 Primary Efficacy Analysis

The primary analysis will be based on the change from baseline in 3-component Mayo Clinic Score (consisting of subscores for stool frequency, rectal bleeding and findings of endoscopy) at Week 12 using an analysis of covariance (ANCOVA) model with terms for treatment, current oral corticosteroid use, prior exposure to TNFα antagonists and baseline value as covariate. Least-squares mean (LSM) by treatment group and its 90% confidence interval (CI), and least-squares mean difference between treatment group and its 90% CI will be reported.

4.3.2 Secondary Efficacy Analyses

The secondary analyses include the following:

- The proportion of patients who achieve endoscopic improvement will be analysed individually using Cochran-Mantel-Haenszel (CMH) test adjusted for the stratification factors of presence or absence of current oral corticosteroid therapy at baseline and previous exposure to TNFα antagonists, to compare the difference of proportions between treatment groups. The CMH stratified risk difference and its 90% confidence interval will be provided.

- The analyses based on the change from baseline in 2-component Mayo Clinic Score and Total Mayo Clinic Score at Week 12 will be performed using an ANCOVA model with terms for treatment, current oral corticosteroid use, prior exposure to TNFα antagonists and baseline value as covariate. Least-squares mean (LSM) by treatment group and its 90% confidence interval (CI), and least-squares mean difference between treatment group and its 90% CI will be reported.

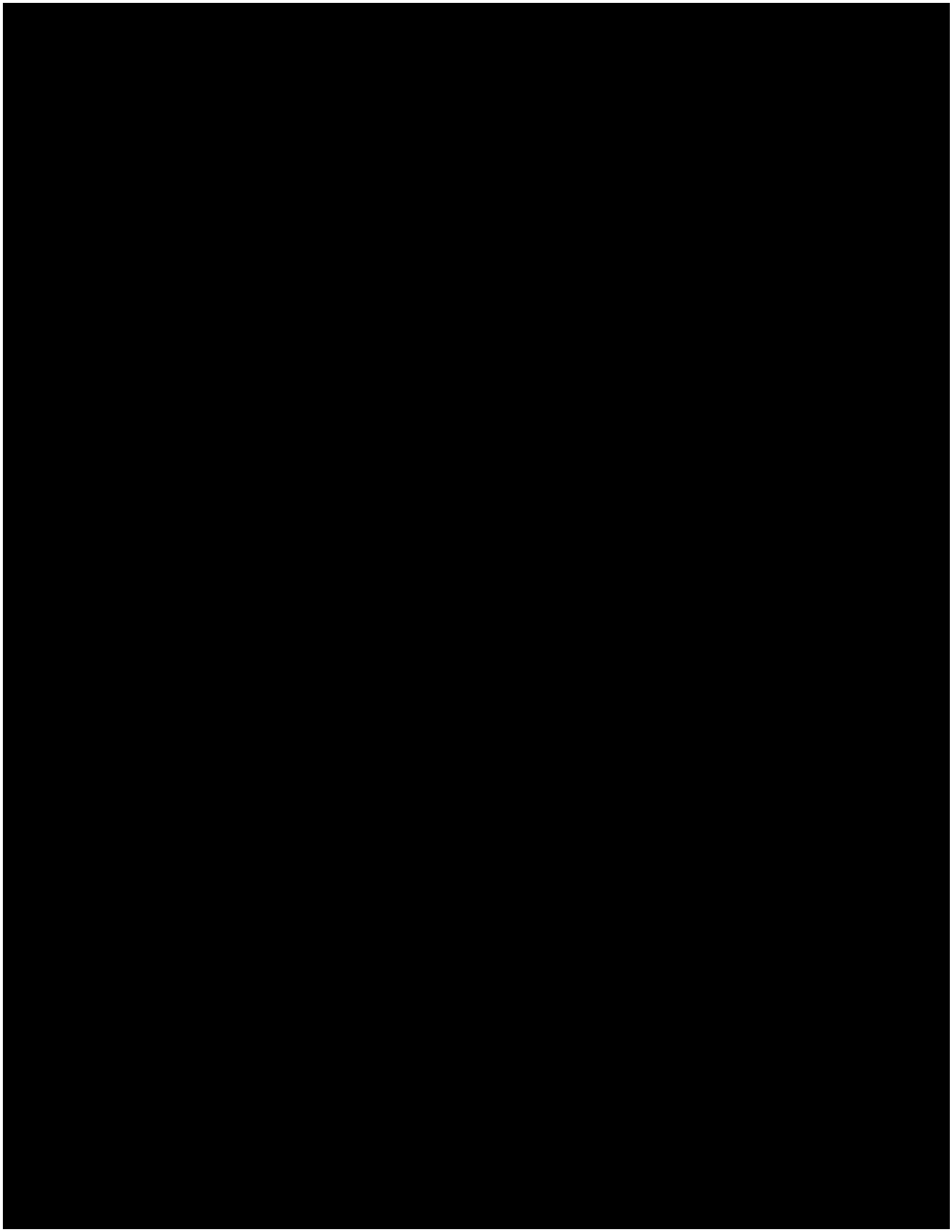
4.3.3 Exploratory Efficacy Analyses

Trichotomous composite endpoint of clinical remission and clinical response at Week 12 is an ordinal categorical endpoint with 3 categories (2, 1, 0). It will be analysed using Cochran-Mantel-Haenszel method adjusted for the stratification factors of presence or absence of current oral corticosteroid therapy at baseline and previous exposure to TNFα antagonists. The testing statistic and p-value will be computed using modified ridit scores for between-treatment comparison.

For proportion based efficacy endpoints such as [redacted] patients who achieve clinical remission and proportion of patients who achieve clinical response similar analyses specified for secondary analyses, such as CMH test will be performed.

[redacted]

Dose response analyses will be performed using an ANCOVA model for continuous variables, or logistic regression model for categorical variables, as appropriate. Treatment group and appropriate baseline covariate will be included in the analysis model. Statistical testing of dose response trend will be based on appropriate contrast statement assuming monotonic dose response profile in placebo (lowest response), APD334 low dose (middle-range response) and APD334 high dose (highest response).



4.5 Definition of Compliance Measure

Compliance will be assessed using patient data recorded in the drug accountability form of the electronic case report forms (eCRFs). On each day, a patient should take his/her assigned treatment. The compliance rate for each patient will be computed as 100% x (actual number of tablets taken over the study period)/ (designated total number of tablets that should have been taken over the study period). Study period is defined as the number of days that the patient has been in the active treatment phase of the trial. Compliance rates will be summarized for each treatment group.

5 SAFETY ANALYSES

Safety and tolerability will be assessed by a review of all safety parameters including adverse experiences (AEs), laboratory safety parameters, vital signs, and electrocardiogram (ECG). Only summary tabulations (N, mean [or median], SD, mean [or median] change/percent change) and 90% CIs for between-group differences will be obtained. Adverse experiences will only be presented as summary tabulations.

5.1 Safety Population

The analyses for all safety outcomes (categorical or continuous measures) will use the safety population which consists of all randomized patients who received at least 1 dose of study drug; in addition, if a patient is found to have taken a study therapy for the entire duration of the study different from that to which he/she was randomized, then the patient is counted in the treatment group of the drug he/she actually received.

For analysis based on laboratory measurements, at least 1 laboratory test post-randomization is required for inclusion in the safety population. When assessing change from baseline, a baseline measurement is also required. No missing data will be imputed for the safety analysis.

5.2 Dose and Duration

The duration of treatment for each patient will be assessed by calculating the number of weeks on drug. For each dosage, the range (minimum and maximum) of values for days on drug and the mean number of days on drug and exposure in patient-weeks will be calculated.

5.3 Adverse Events

Adverse events will be coded using the most current Medical Dictionary for Drug Regulatory Affairs (MedDRA, version 18.0 or later) and tabulated, including categorical information of interest such as onset and resolution times, time of onset relative to dose, severity at onset, maximum severity, causal relationship to study medication, and action taken. AEs will be regarded as 'pre-treatment' if they occur between screening and the time of administration of the first dose of APD334 or placebo and will be recorded as medical history. All other AEs that occur after the first dose of study medication will be considered to be 'treatment-emergent'.

Treatment-emergent AEs (TEAE) will be listed by patient and by treatment. They will be summarized per treatment and expressed in terms of maximum severity and relationship to study medication. The incidence of TEAEs classified according to system organ class (SOC) will be summarized by treatment group. TEAEs will also be summarized by maximum intensity (assessed according to the Common Terminology Criteria for Adverse Events v4.03[8]definitions) and relatedness to study medication.

Summaries of the number (%) of patients in each treatment group with at least 1 TEAE, classified according to MedDRA system organ class and preferred term, will also be provided for:

- Drug-related TEAE
- Treatment-emergent AEs leading to permanent discontinuation of study medication (study medication discontinued or withdrawal from study).
- Serious adverse events (SAEs)

Serious adverse events will be listed by patient and by treatment. If there are no SAEs at the end of the study, the tables or listings will state that there are no SAEs in the study.

5.4 Physical Examinations

Physical examination dates will be listed.

5.5 Concomitant Medication

Pre-treatment and concomitant medication administered during the study will be listed. Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHODRUG Dictionary) and will be tabulated by drug class and preferred term.

5.6 Vital Signs

Individual vital sign measurements will be listed by treatment and summarized using descriptive statistics. Summary statistics will also be provided for change from baseline in vital sign measurements by treatment. If Day 1 pre-dose assessment is completed then this will be used for baseline. If the Day 1 pre-dose assessment is missing, then the last non-missing assessment prior to randomization will be used.

5.7 Clinical Laboratory Values

Individual lab values will be listed by treatment and visit, and summarized using descriptive statistics. Summary statistics will also be provided for change from baseline in lab values. Shift tables from baseline to last double-blind visit will also be produced for the laboratory assessments based on the categories of Low, Normal, and High. Baseline is defined as the last value of a specific endpoint measured before first dose of study medication.

If any laboratory value falls above or below the upper or lower level of quantification, the value of the upper or lower level of quantification will be taken (e.g. <0.2 will become 0.2) for summaries but left as recorded in the listing.

5.8 Safety ECG (12-lead ECG) and Holter Monitoring

[redacted]

6 OTHER ANALYSES

6.1 Interim Analyses

Sponsor may plan an unblinded interim analysis at time when sufficient number of patients have finished 12 weeks of treatment. If sponsor decides to perform an interim analysis, it will be operated by an independent Data Monitoring Committee (DMC).

6.2 Subgroup Analyses

Subgroup analyses for the primary efficacy endpoint (i.e., improvement of 3-component Mayo Clinic Score consisting of subscores for stool frequency, rectal bleeding and findings of endoscopy at Week 12) and other important efficacy endpoints will be performed in order to explore whether the treatment effects are consistent across different subgroups. The baseline patient characteristics below are the subgroup factors to be explored.

- Sex (Male, Female)
- Age: > or ≤ median age, ≥ or < 65 years
- Race
- Presence or absence of current oral corticosteroid usage
- Previous exposure to TNFα antagonists
- Responders vs. subjects who had inadequate / loss of response in previous exposure to TNFα antagonists

[redacted]

- Baseline Total Mayo score ≤ 8 vs > 8.

The change from baseline in 3-component Mayo Clinic Score at Week 12 using an ANCOVA model with terms for treatment, subgroup, treatment-by-subgroup interaction, and baseline value as covariate will be provided. Least-squares mean (LSM) by treatment group for each of the subgroups and its 90% confidence interval (CI), and least-squares mean difference between treatment group for each of the subgroups and its 90% CI will be reported. A forest plot for all subgroups will be presented for the least-squares mean differences.

In addition to the subgroup analyses for the primary endpoint, the following analyses will be performed:

[redacted]

6.3 Multiplicity

Primary comparison is between APD334 2 mg vs placebo [redacted]

The primary analysis will compare PMS#1 change from baseline between treatment groups using a one-sided test at the 0.05 level of significance and 90% CI will be reported. If the result is significant, the primary hypothesis will be considered satisfied and this study will be declared positive. If this primary test is significant the secondary testing will proceed as following.

Secondary endpoints:

1. Proportion of patients who achieve endoscopic improvement at Week 12
2. Improvement in 2 component Mayo score (score ranging from 0 to 6, including rectal bleeding and findings on endoscopy) at Week 12
3. Improvement in Total Mayo Clinic Score (score ranging from 0 to 12, including stool frequency, rectal bleeding and findings on endoscopy, and physician's global assessment) at Week 12.

If the primary testing is significant, testing for secondary endpoints will proceed in pre-specified order. Provided that all previously specified tests are statistically significant ($p < 0.05$), each of these comparisons will be assessed using the one-sided test at the 0.05 level of significance. However, if a comparison is not statistically significant, then subsequent comparison(s) will be considered exploratory.

Exploratory endpoints:

- The trichotomous composite endpoint of clinical remission and clinical response (score ranging 0 to 2: score 2 for achieving both clinical remission and clinical response; 1 for achieving only clinical response, and 0 for achieving neither) at Week 12.
- The proportion of patients achieving clinical remission at Week 12
- Proportion of patients who achieve clinical response at Week 12

[redacted]

Non-formal statistical testing on treatment effect will be performed and reported along with nominal p-values.

6.4 Statistical Assumptions Checking

For score or continuous measures, the assumptions for ANCOVA may be checked by the following tests:

- parallel regressions tested in a supplemental ANCOVA that includes baseline by treatment interaction term
- assumption of homogeneous variability between groups may be tested, using Levene's test or the Brown-Forsythe test, whichever appropriate
- visual inspection of linearity of regression

If there is evidence of gross violation of the assumptions for ANCOVA, then the offending endpoint may be analysed by ordinal logistic regression.

For categorical measures, if a logistic model gives non-estimable odds across a treatment group due to complete or quasi-separation then the variable causing the separation may be removed from the model. This may be noted in the results table.

If the proportional odds assumption in an ordinal logistic regression model is violated, which is indicated by p-value<0.05 by score test for the proportional odds assumption, then treatment effects may be estimated using analysis of covariance (ANCOVA).

6.5 Data/Study Reviewing Committees

The Data Safety Monitoring Board (DSMB) will monitor emerging study safety data and has the responsibility to review specific safety data reports, and to request additional reports as needed. Unblinded data reports generated by the unblinded statistician will be reviewed for adverse treatment effects and patient safety. The DSMB must keep results from blinded and unblinded data reports confidential. The roles and responsibilities of the DSMB will be outlined in a separate charter.

7 STATISTICAL TECHNICAL ISSUES

7.1 Planned Statistical Power and Sample Size

For the comparison of the primary efficacy measure in PMS#1 (including subscores for stool frequency, rectal bleeding and findings of flexible proctosigmoidoscopy) change from baseline, a sample of ~39 evaluable patients (having non-missing PMS at baseline and at Week 12) per group will provide approximately 80% power to detect a difference of 1.15 at α=5% (1-sided test) based on an estimated pooled standard deviation 2.03. Sponsor will enroll sufficient patients to reach at least 117 evaluable patients.

7.2 Method of Assigning Participants to Treatment Groups Study

Patients who meet all the entry criteria and are eligible for the study will report to the investigator to be randomized on Day 1. Sites will randomize eligible patients for entry into the study in a 1:1:1 ratio to receive once daily doses of 1 mg APD334, 2 mg APD334, or matching placebo across the entire study, and is not constrained within a center. Randomization will be stratified by presence or absence of current oral corticosteroid usage and previous exposure to TNFα antagonists. All pre-specified efficacy analysis model will include for these two factors using CRF recorded data. Potential discrepancy between CRF recorded data and IVRS data entered by site staff will be investigated. A post-hoc analysis may be performed to evaluate the impact on main study endpoints caused by such discrepancy.

7.3 Blinding/Unblinding

The sponsor, patients, and personnel involved with the conduct of the study, with the exception of the clinical supply staff, safety staff, and the unblinded statistician supporting the Data Safety Monitoring Board (DSMB), will be blinded to the identity of study medication until the database lock. Blinding is accomplished by the random, masked, assignment of allocation numbers to the treatment groups, and by ensuring the drug supplies administered in the treatment groups appear identical. The contract research organization (CRO) will obtain written consent from Arena prior to breaking the code.

Breaking of the randomization code without Arena permission is expressly forbidden except in the event of a medical emergency where the identity of the study medication must be known in order to properly treat the patient. In the event of a medical emergency, it is requested that the investigator make every effort to contact the study monitor or designee prior to breaking the code. If the blind is broken, the individual responsible should document the date, time, and reason for breaking the blind. A written communication should be sent to Arena within 1 working day.

7.4 Handling of Missing Data and Sensitivity Analysis

7.4.1 Multiple Imputation for Primary Efficacy Analysis

Multiple imputation has 3 distinct phases: Imputation, Analysis and Combining. During the Imputation Phase, the missing data values are filled-in to form a complete dataset. This is done

'm' times. For the Analysis Phase, each of the 'm' complete datasets are analysed by a statistical model (as specified in Section 4.3 for each endpoint). The Combining Phase pools the parameter estimates with their standard errors across the 'm' complete datasets and produces an average estimate with a standard error calculated using Rubin's formula[9,10] this allows for the uncertainty between imputations as well as the variability within each analysis.

Multiple imputation assumes that the data are missing at random (MAR). MAR assumes that the missingness does not depend on the actual missing values, but that the missing data can be completely explained by the observed data. To test the robustness of the multiple imputation strategy a tipping point analysis may be implemented.

A tipping point analysis may be implemented for the primary endpoint to explore the influence of missing data on the overall conclusion of the primary efficacy analysis results. In this analysis, a wide spectrum of assumptions can be imposed regarding the missing data (from less conservative to more conservative) to find the "tipping" point assumption, at which conclusions change from being favorable toward APD334 to being unfavorable. After such a tipping point is determined, clinical judgment can be applied as to the plausibility of the assumptions underlying this tipping point. A delta (δ) method is often used to clearly formulate clinical assumption about Missing Not At Random (MNAR) mechanism. More specifically, the assumption is that, after dropout, subjects from APD334 have, on average, their PMS score worsen by some amount δ (δ =0 to 9, with a max PMS score of 9) from the time of dropout to Week 12.

Whether the missing data pattern is monotone or arbitrary (non-monotone), a fully conditional specification (FCS) method[11,12] with predictive mean matching for continuous variables[13,14] can be used. The FCS method allows for separate conditional distributions for each imputed variable. The predictive mean matching approach creates a regression model using parameters sampled from the posterior distribution and then a predicted value for each missing value is computed. The missing value is replaced by randomly selecting an observation from a set of 'k' values that are the closest predicted values to the missing predicted value.

In the case of Mayo scores, a multiple imputation approach is implemented for the missing individual components of the Total Mayo Score (TMS), which include stool frequency, rectal bleeding, findings of flexible proctosigmoidoscopy (endoscopy), and physician global assessment. Each individual subscore is imputed and then the composite endpoint measures are created (see list of endpoints below) following the specified algorithms or formulas in the SAP. The rectal bleeding, stool frequency, endoscopy, physician's global assessment, treatment, stratification factors (oral corticosteroid use and prior exposure to TNFα antagonists), biomarker variables (e.g. [redacted] inform the imputation model. The imputation model may include all these variables at all planned time points (e.g. Screening, Baseline, Week 1, 2, 4, 8, 12 and may be specific for each variable). These variables are included because they are correlated with the endpoint and/or missingness. The multiple imputation process creates [redacted] complete datasets using PROC MI in SAS 9.4. During post-processing, the endpoints are derived in each of the [redacted] complete datasets. The endpoints and analysis model in parentheses are listed below:

- PMS#1 (bleeding, stool, endoscopy) change at Week12 (ANCOVA)

- %-endoscopic improvement at W12 (Logistic regression)
- 2-component Mayo scores (bleeding, endoscopy) change at Week12 (ANCOVA)
- Total Mayo score (4-component) change at Week12 (ANCOVA)
- %-clinical remission at Week12 (Logistic regression)
- %-clinical response at Week12 (Logistic regression)
- Endoscopic score change at Week12 (ANCOVA)
- Combination of clinical remission and response (trichotomy) at Week12 (Ordinal Logistic)

An analysis model (e.g. ANCOVA or Logistic as specified in the Section 4.3) for each endpoint is run on each of the multiply imputed datasets. Thus, for any estimator of interest, there will be estimates with standard errors obtained; these are pooled using PROC MIANALYZE to produce an overall estimate and a standard error (based on Rubin's formula) with associated confidence interval and p-value.

The multiple imputation model will be pre-defined before database lock. However, there could be certain adjustments due to unexpected data issues after unblinding treatment and . All post-unblinding modifications to the multiple imputation model or approaches to address missing data will be described in the Clinical Study Report.

7.4.2 Additional Sensitivity Analyses

8 DATA HANDLING CONVENTIONS

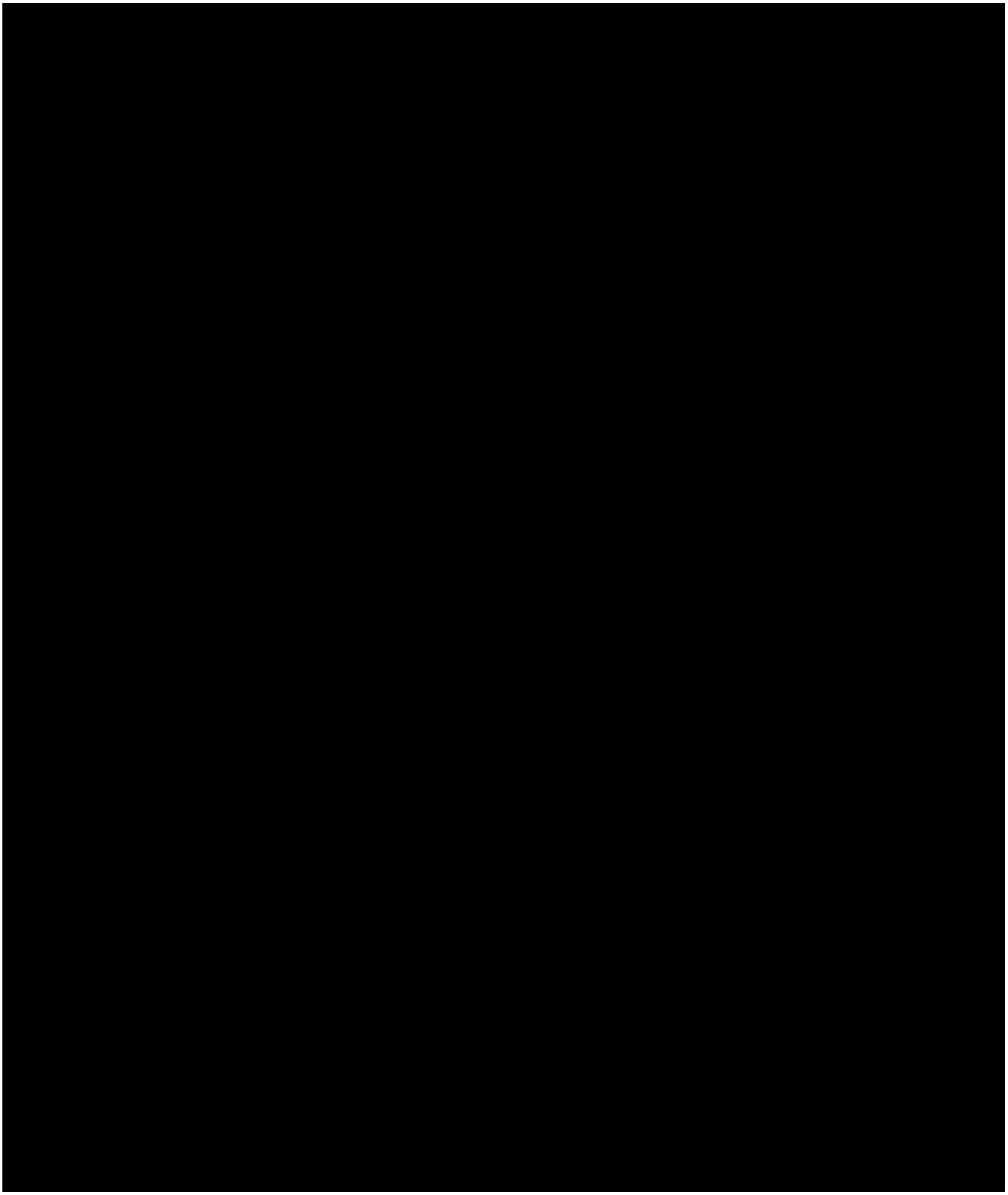
8.1 Baseline Definitions or Conventions

For the stool frequency and rectal bleeding components of the Mayo Clinic Score, baseline will be based on patient diary entries within the 10 days prior to randomization. The scores from the 3 most recent days prior to the actual randomization day will be averaged and rounded to the nearest integer. The rounding will be applied to each subscore prior to the creation of the total score. Note that the day prior, day of and day after proctosigmoidoscopy cannot be used for patient diary entry because of the required bowel prep for the procedure. Patients who have less than 3 days of diary data during screening are not eligible for randomization. Baseline endoscopy subscore will be based on flexible proctosigmoidoscopy results within 10 days prior to randomization.

For other efficacy and safety endpoints, baseline value is defined as the last pre-randomization measurement, unless otherwise specified.

8.2 Time Points, and Day Ranges

Since it is not always possible for all study participants to come in for their clinic visits on the exact day specified in the protocol schedule, the "Week" of a patient's visit will be defined by the following relative day ranges. Tables 3 to 7 below give the mapping of relative day ranges to Week.



If a patient has more than one assessment in a window, the assessment date closest to the target date will be selected. If a patient has 2 assessments that are equidistant from the target date, the later assessment will be selected.

8.3 Description of Data Handling Procedures Prior to Unblinding

All data will be screened, reviewed, and declared clean before data are unblinded according to PPD (Data Management CRO) guidelines and standard operating procedures. The PPD unblinded database will be locked in order to insure that analyses in response to regulatory queries are performed on the same data used for submission. The freeze and unblinding will occur at the end of the study.

9 CHANGES IN THE PLANNED ANALYSIS FROM THE PROTOCOL

The following secondary objective and endpoint have been added to the SAP in comparison with Protocol Amendment #6:

1. 2-Component Mayo Clinic Score

The following changes to the secondary objectives and endpoints have been made:

1. The order has been changed to: Endoscopic Improvement, 2-Component Mayo Score and Total Mayo Score

2. Other objectives/endpoints previously listed as secondary (i.e. Trichotomous score, Clinical remission and Clinical response) are considered to be exploratory

The following exploratory objectives and endpoints have been added to the SAP in comparison with Protocol Amendment #6:

In addition, a change to the sample size calculation in comparison with Protocol Amendment #6 has been made:

Rationale for sample size calculation change:

In Protocol Amendment 06, the primary objective was changed to reflect a broad examination of potential clinical improvement across three domains (stool frequency, rectal bleeding, and endoscopic results), which is appropriate for a proof-of-concept study. The current endpoint is more sensitive to detecting clinically important differences between two doses of APD334 and placebo compared to a dichotomous complete remission endpoint.

The 95% CIs have been changed to 90% CIs for all analyses in comparison with Protocol Amendment #6.

The clinical remission definition has been updated to include “a rectal bleeding score of 0 or 1” in comparison with Protocol Amendment #6.

The clinical response definition has been updated as "a patient has achieved clinical response if he/she meets criteria of clinical remission, or meet criteria of clinical response defined in the protocol."

The following change to the primary efficacy analysis has been made:

A Phase II POC study does not have as strict criteria as a pivotal Phase III study. In recent years regulators have been critical of last observation carried forward (LOCF) and similar approaches for handling missing data and have advocated using more scientifically justifiable (e.g. multiple imputation methods) in primary analyses of Phase III studies alongside testing assumptions. Recommendation 10 of The Prevention and Treatment of Missing Data in Clinical Trials[15] states, "Single imputation methods like last observation carried forward and baseline observation carried forward should not be used as the primary approach to the treatment of missing data unless the assumptions that underlie them are scientifically justified". The EMA Guideline on Missing Data in Confirmatory Clinical Trials[16] states, "The risk of underestimating the variance of treatment effect when imputing can be reduced by proper implementation of techniques such as multiple imputation."

For APD334-003 Phase II POC study, approximately 10% missing data is anticipated for the endoscopic component of the 3-component Mayo Clinic Score at Week 12. With 10% missing data, the balance in the original randomisation (underpinning the Intention-To-Treat (ITT) principle) could be lost which could result in biased estimates. Multiple Imputation (MI) can reduce bias arising from missing data.

Therefore, multiple imputation will be used to handle missing data in the analysis of the for efficacy endpoints at Week 12 for APD334-003 study. The original planned efficacy analysis in the protocol therefore becomes a sensitivity analysis in support of the new primary efficacy analysis.

10 REFERENCES

1. APD Protocol 003: A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Parallel Group, Multi-Center Study to Investigate the Safety and Efficacy of APD334 in Patients with Moderately to Severely Active Ulcerative Colitis.

2. Kornbluth A, Sachar DB, MACG and The Practice Parameters Committee of the American College of Gastroenterology. Ulcerative Colitis Practice Guidelines in Adults: American College of Gastroenterology, Practice Parameters Committee. Am J Gastroenterol 2010;105:501-523; doi:10.1038 / ajg.2009.727 https://www.nature.com/articles/ajg2009727

3. Crohn's and Colitis Foundation. Treatment Options (patient brochure) http://www.crohnscolitisfoundation.org/what-are-crohns-and-colitis/newly-diagnosed/treatments.pdf

4. Scoville EA, David A. Schwartz DA. Endoscopy in inflammatory bowel disease: advances in disease management. Gastrointestinal Endoscopy 2017;86:952-961; https://doi.org/10.1016/j.gie.2017.08.034 http://www.giejournal.org/article/S0016-5107(17)32243-5/fulltext

5. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Ulcerative Colitis: Clinical Trial Endpoints Guidance for Industry (DRAFT GUIDANCE), August 2016. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM515143.pdf

6. European Medicines Agency. Guideline on the development of new medicinal products for the treatment of ulcerative colitis (Draft). 21 July 2016. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/07/WC500211431.pdf

7. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. New England Journal of Medicine 1987; 317:1625-1629; DOI: 10.1056/NEJM198712243172603 http://www.nejm.org/doi/full/10.1056/NEJM198712243172603

8. Common Terminology Criteria for Adverse Events v4.03 (CTCAE). Publish Date: June 14, 2010. Accessed October 2010. http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

9. Rubin, D.B. (1987), Multiple Imputation for Nonresponse in Surveys, New York: John Wiley & Sons, Inc.

10. Little, R.J.A. and Rubin, D.B. (2002), Statistical Analysis with Missing Data, New York: John Wiley & Sons, Inc.

11. Brand, J. P. L. (1999), "Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets," Ph.D. dissertation, Erasmus University, Rotterdam.

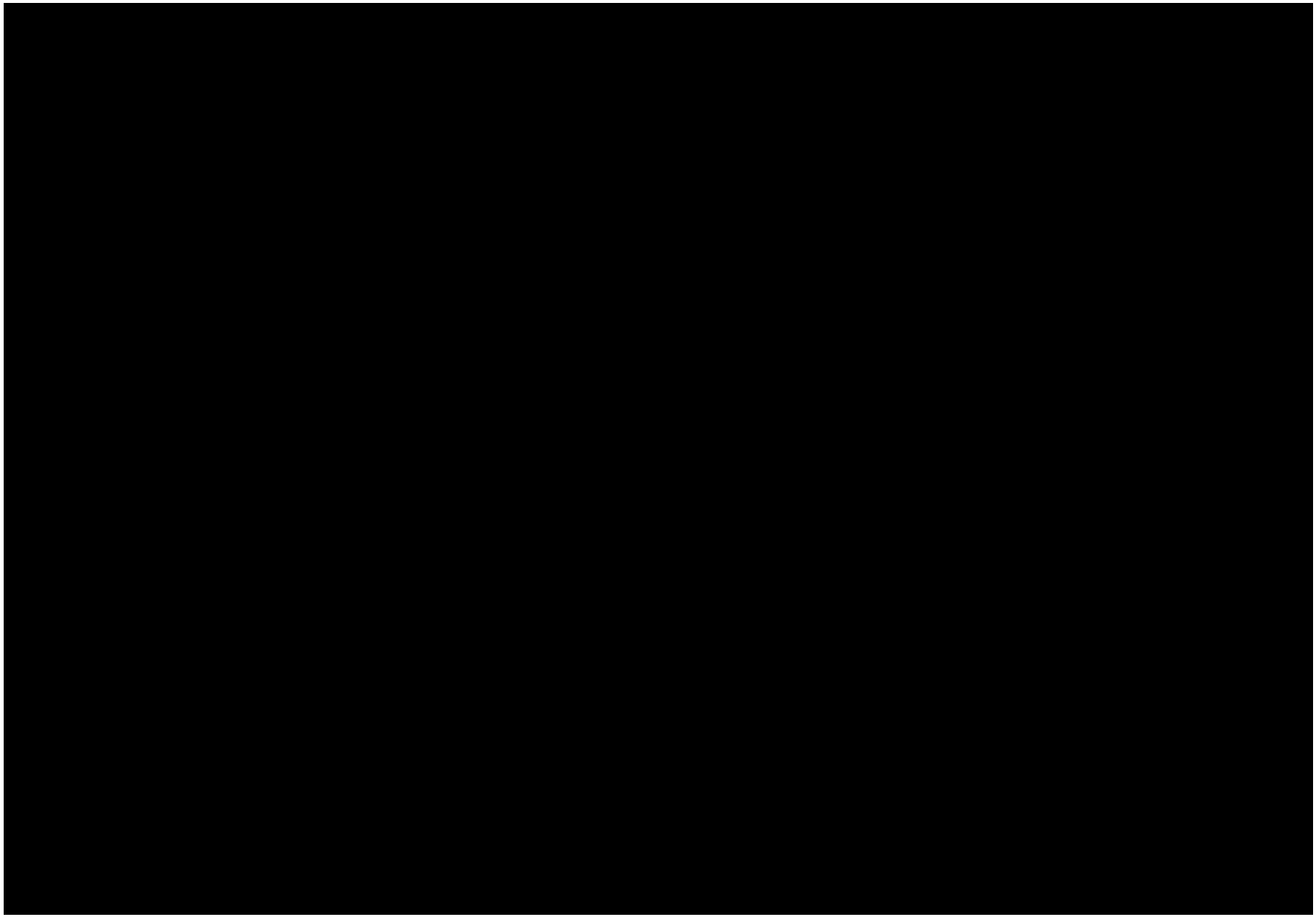
12. van Buuren, S. (2007), "Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification," Statistical Methods in Medical Research, 16, 219–242.

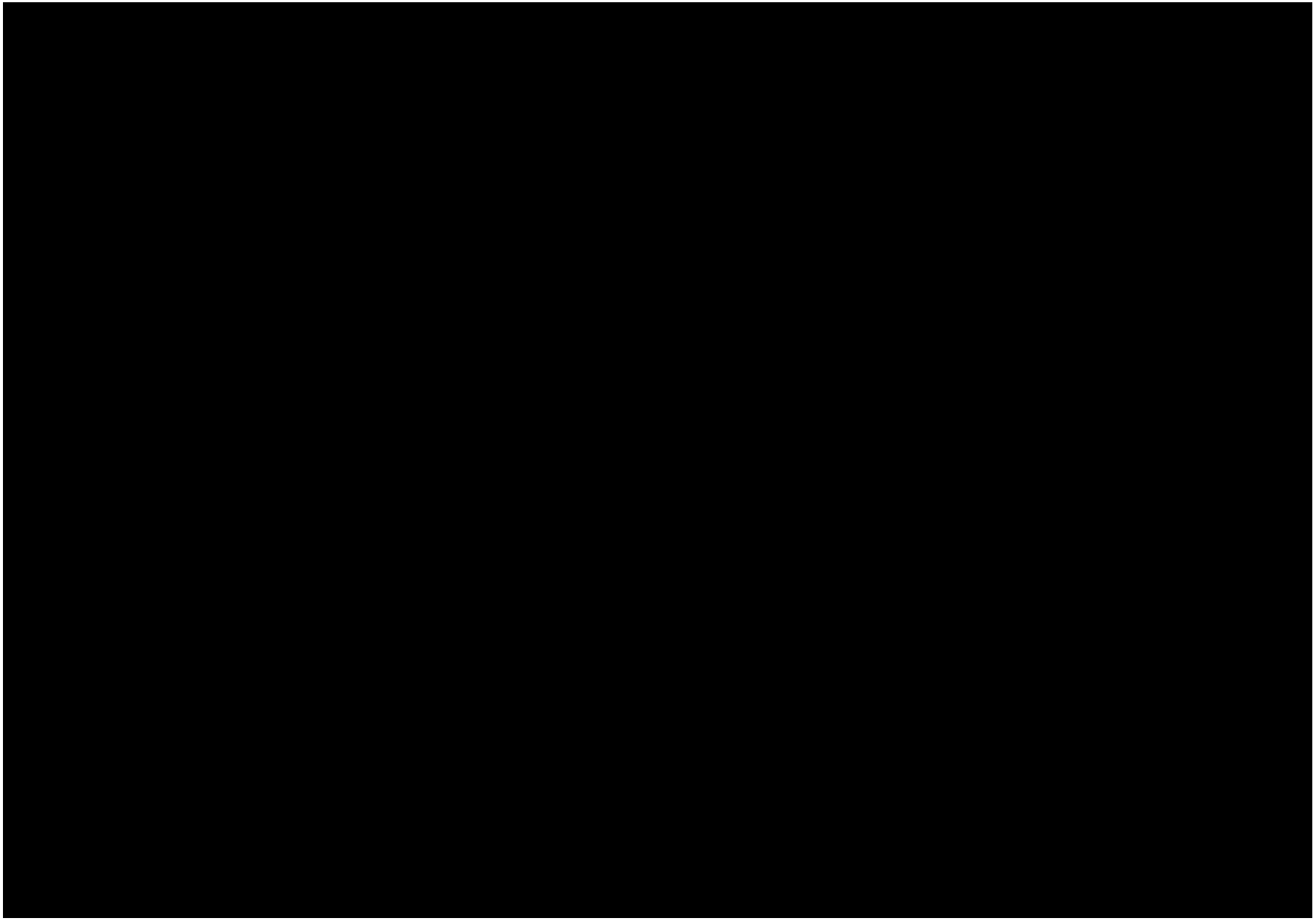
13. Heitjan, F. and Little, R. J. A. (1991), "Multiple Imputation for the Fatal Accident Reporting System," Applied Statistics, 40, 13–29.

14. Schenker, N. and Taylor, J. M. G. (1996), "Partially Parametric Techniques for Multiple Imputation," Computational Statistics and Data Analysis, 22, 425–446.

15. National Research Council. (2010). The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

16. Committee for Medical Products for Human Use. Guideline on missing data in confirmatory clinical trials. London: European Medicines Evaluation Agency, April 2009 (http://www.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf).





Appendix 2 Listing of Efficacy Parameters and Study Populations Analysed

Efficacy Parameters	Response Analysed	Study Populations Analysed
Primary Endpoint		
3-Component Mayo Score[1]	Change from Baseline	MITT, CP, ITT
Secondary Endpoints		
Endoscopic Improvement	Proportion	MITT, CP, ITT
2-Component Mayo Score	Change from Baseline	MITT, CP, ITT
Total Mayo Score	Change from Baseline	MITT, CP, ITT
Exploratory Endpoints		
Trichotomous Composite Score	Ordinal Categorical	MITT, CP, ITT
Clinical Remission	Proportion	MITT, CP, ITT
Clinical Response	Proportion	MITT, CP, ITT
[redacted]		

[1] includes subscores for stool frequency, rectal bleeding and findings on endoscopy

[redacted]

Appendix 3 Mayo Clinic Score

Mayo Scoring System for Assessment of Ulcerative Colitis Activity[3]

Stool Frequency†

0 = Normal number of stools for this patient
1 = 1–2 stools more than normal
2 = 3–4 stools more than normal
3 = 5 or more stools more than normal
Subscore, 0 to 3

Rectal bleeding‡

0 = No blood seen
1 = Streaks of blood with stool less than half the time
2 = Obvious blood with stool most of the time
3 = Blood alone passes
Subscore, 0 to 3

Findings on endoscopy†

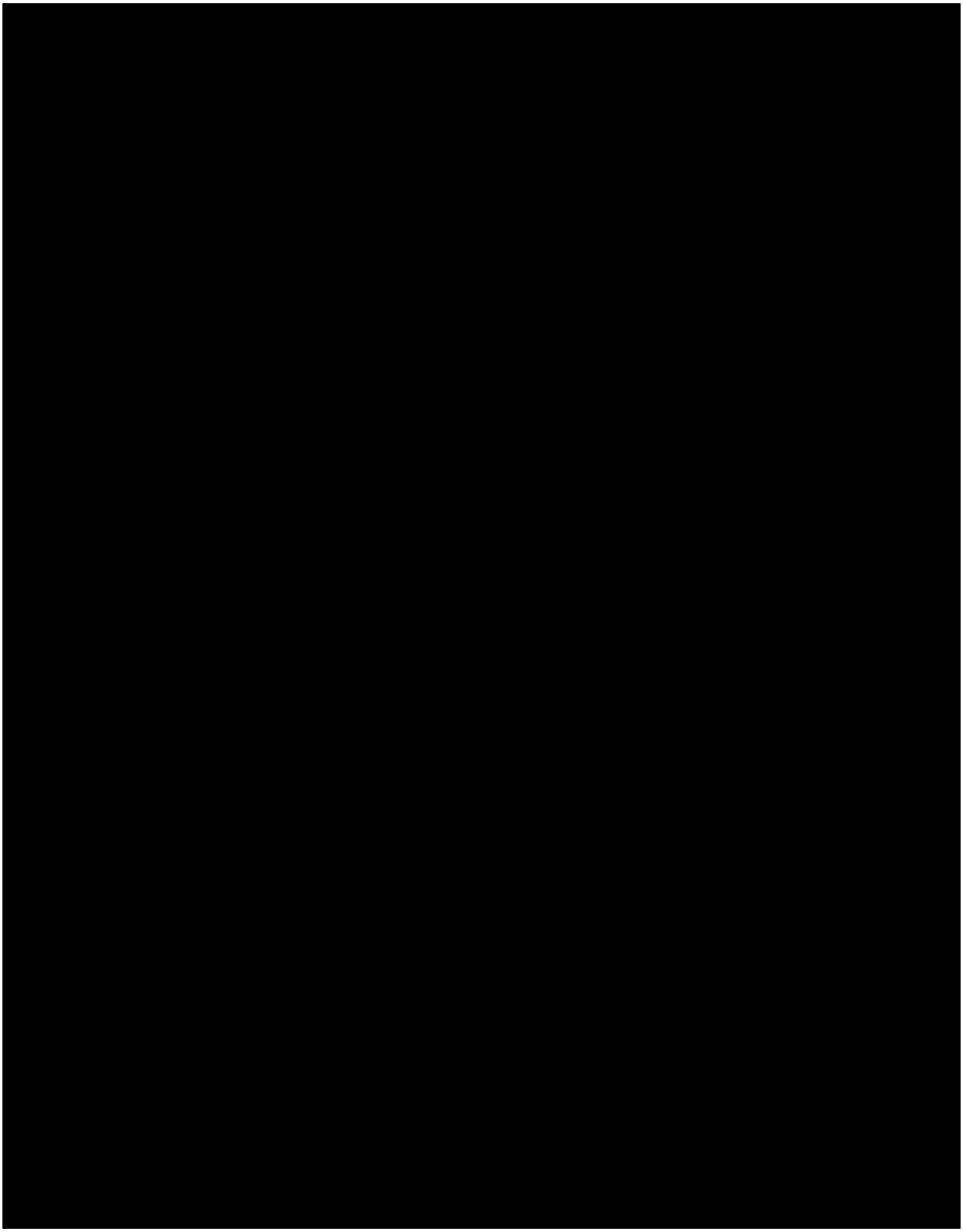
0 = Normal or inactive disease
1 = Mild disease (erythema, decreased vascular pattern)
2 = Moderate disease (marked erythema, lack of vascular pattern, friability, erosions)
3 = Severe disease (spontaneous bleeding, ulceration)
Subscore, 0 to 3

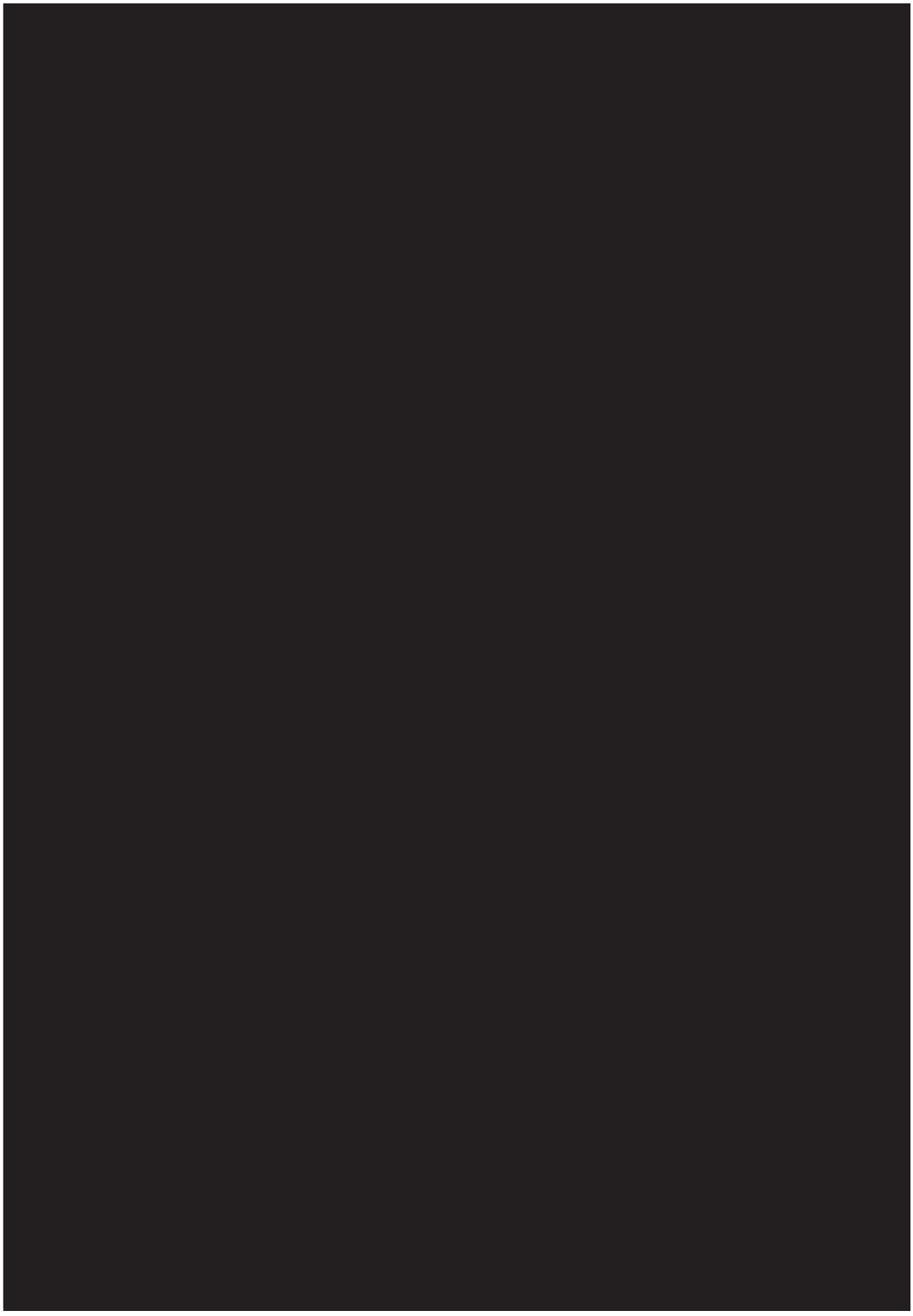
Physician's Global Assessment§

0 = Normal
1 = Mild disease
2 = Moderate disease
3 = Severe disease
Subscore, 0 to 3

† Each patient serves as his or her own control to establish the degree of abnormality of the stool frequency.
† The daily bleeding score represents the most severe bleeding of the day.
§ The physician's global assessment acknowledges the three other criteria, the patient's daily recollection of abdominal discomfort and general sense of well-being, and other observations, such as physical findings and the patient's performance status.

Note: The Mayo score ranges from 0 to 12, with higher scores indicating more severe disease.[4]





Appendix 5 Multiple Imputation Model