1.0 Title Page

Statistical Analysis Plan

Study M10-870

A Multi-Center, Open-Label Study of the Human Anti-TNF Monoclonal Antibody Adalimumab to Evaluate Long-Term Safety and Tolerability of Repeated Administration of Adalimumab in Pediatric Subjects with Ulcerative Colitis Who Completed the Study M11-290

Date: 08 Aug 2019

Version 1.0

2.0 Table of Contents

List of Tables

3.0 Introduction

This statistical analysis plan (SAP) will provide details to further elaborate statistical methods as outlined in the protocol of Study M10-870, "A Multi-Center, Open-Label Study of the Human Anti-TNF Monoclonal Antibody Adalimumab to Evaluate Long-Term Safety and Tolerability of Repeated Administration of Adalimumab in Pediatric Subjects with Ulcerative Colitis Who Completed the Study M11-290" and describe analysis conventions to guide the statistical programming work. The analysis plan is based on the study protocol Amendment 3 dated 30 July 2019.

The analyses for the Japan Sub-Study will be described in a separate SAP. All described analysis in this SAP will be for subjects enrolled outside of Japan.

The statistical analysis will be performed using the SAS® software version 9.2 or later.

4.0 Study Objectives, Design and Procedures

4.1 Objectives

The objective of the study is to evaluate the long-term safety, tolerability, and maintenance of clinical response, of repeated administration of adalimumab (ADA) administered subcutaneously (SC) in pediatric subjects with ulcerative colitis (UC) who participated in, and successfully completed, Protocol Study M11-290 through Week 52.

4.2 Design Diagram

This study is a Phase 3, multi-center, open-label study designed to evaluate the long-term maintenance of clinical response, safety and tolerability of ADA in pediatric subjects with ulcerative colitis.

Subjects are eligible to enroll in the Study M10-870 if they have successfully completed Protocol Study M11-290 through Week 52 and meet all of the inclusion criteria and none of the exclusion criteria for Study M10-870.

The Week 52 visit from the Study M11-290 serves as the Baseline Visit for subjects entering Study M10-870.

All subjects receive open-label therapy as follows beginning at the Baseline Visit in Study M10-870:

- Subjects who enroll into the study from blinded treatment in Study M11-290 receive open label 0.6 mg/kg (maximum of 40 mg) every other week (EOW) of adalimumab
- Subjects who received open label 0.6 mg/kg (maximum of 40 mg) every week (EW) of adalimumab in Study M11-290 maintain the same dose in Study M10-870

Subjects who demonstrate a disease flare may have their dose escalated. Subjects who have been in remission (PMS ≤ 2 with no individual subscore > 1) for at least 8 weeks and for at least 2 consecutive visits may have their dosage de-escalated. Subjects who demonstrate disease flare after dose de-escalation have an opportunity to re escalate their dose back to adalimumab every week (EW) dosing.

Subjects with persistently uncontrolled disease while on adalimumab 40 mg EW (maximum dose) are to be discontinued from the study.

The duration of the study is up to 298 weeks (approximately 5.5 years).

More details are described in the protocol, including the Study Activities table.

4.3 Sample Size

Subjects who successfully completed Study M11-290 through Week 52 are potentially eligible to participate in this study. Originally, based on an expected number of 155 subjects being re-randomized at Week 8 of the parent study, an expected Study M11-290 completion rate of 66% and a roll-over rate of 80% of eligible subjects (based on previous experience with the pediatric CD Studies M06-806 and M06-807) it was expected that approximately 85 subjects from Study M11-290 would enroll into the

main Study M10-870 study and 8 subjects from the Japanese sub-study. However, as per Study M11-290 protocol Amendment 5 the parent study will be closed with N = 101 (N = 93 in main study + N = 8 in Japan sub-study) and the expected sample size in Study M10-870 is approximately 60 subjects (55 Ex-Japan and 5 in Japan).

4.4 Interim Analysis

An interim analysis on safety and efficacy will be performed and reported in a full Interim Clinical Study Report for inclusion of the Study M10-870 data in the submission of applications for regulatory approval of the pediatric UC indication in conjunction with the Study M11-290 results. Additional interim analyses may be performed, as needed.

5.0 Analysis Populations

5.1 Definition for Analysis Populations

The Full Analysis Set will be used for all analyses and includes all subjects who received at least one dose of ADA in Study M10-870.

6.0 Analysis Conventions

General Considerations

All statistical analysis results will be reported using descriptive statistics. Descriptive summary statistics will be provided for the demographic and Baseline characteristics, efficacy, and safety parameters.

Continuous variables will be summarized by sample size (N), mean, standard deviation, minimum, Q1, median, Q3, and maximum. Frequency and percentages will be provided for the categorical variables. For dichotomized efficacy variables, data will be summarized by number and percent of subjects with 95% confidence interval.

Week 0 and Baseline Visit Date

Week 0 is defined as the Week 52 visit from the Study M11-290. This date is also referred to as the baseline visit date.

Definition of Baseline Measurement

Unless otherwise notes, the baseline value for a variable is defined as the last non-missing value before the baseline visit date (i.e., at or before the Week 52 visit from the Study M11-290). In some instances Study M11-290-Baseline will be used, but this will always be specified accordingly.

Definition of Final Measurement

Final evaluation during the study: the last non-missing observation collected in the study after the first dose of study drug.

Final evaluation during ADA exposure in Study M10-870 (for laboratory analyses): the last non-missing observation collected in the study within the ADA exposure window in Study M10-870 (see Section 9.1).

Definition of Rx Days (Days Relative to the First Dose of Study Drug)

Rx Days are calculated for each time point relative to the date of first dose of study drug in Study M10-870. They are defined as the number of days between the day of the first dose of study drug and the specific time point. Rx Days are negative values when the time point of interest is prior to the first study drug dose day. Rx Days are positive values when the time point of interest is after the first study drug dose day. The day of the first dose of study drug is defined as Rx Day 1, while the day prior to the first study drug dose is defined as Rx Day –1 (there is no Rx Day 0).

Definition of Analysis Windows

All time points and corresponding time windows are defined based on Rx days. The protocol specified visits and corresponding time windows are presented in the table below. Measures will only be displayed at the visits where they were planned to be measured. For example, Mayo score is collected at Weeks 0, 48, 96, 192 and 288/Premature Discontinuation (dependent on availability of optional endoscopy data at these timepoints).

Assigned Visit	Nominal Day	Time Window (Rx Day Range)
Week 0	1	≤ 1
Week 4	29	2 to 43
Week 8	57	44 to 71
Week 12	85	71 to 127
Week 24	169	128 to 211
Week 36	253	212 to 295
Week 48	337	296 to 379
Week 60	421	380 to 463
Week 72	505	464 to 547
Week 84	589	548 to 631
Week 96	673	632 to 715
Week 108	757	716 to 799
Week 120	841	800 to 925
Week 144	1009	926 to 1093
Week 168	1177	1094 to 1261
Week 192	1345	1262 to 1429
Week 216	1513	1430 to 1597
Week 240	1681	1598 to 1765
Week 264	1849	1766 to 1933
Week 288	2017	1934 to 2101

If more than one assessment is included in a time window the assessment closest to the nominal day should be used. If there were more than one observation with equal distance to the nominal day the latest one will be used in analyses. For any given variable, if more

than one measurement exists for a subject on the same day, the average of the measurements will be considered to be that subject's measurement for that day.

Missing Data Handling

In general, missing Baseline and safety data will not be imputed.

Baseline Value is Missing:

Subjects will be excluded from analysis of change from Baseline if Baseline evaluation is missing.

Missing Efficacy and Outcome Evaluations:

The following methods will be used for imputing missing values to perform the efficacy analyses.

For interim analyses, missing values will only be imputed up to the timepoint a patient could have potentially reached at the data cutoff for analyses.

Observed Cases (OC): missing values will not be used for the observed case analysis.

Last Observation Carried Forward (LOCF): For categorical and continuous efficacy variables the following rules will be used for LOCF approach:

- Baseline and pre-baseline values will not be used to impute the missing post-baseline values, and
- Missing values after baseline will be imputed using the last non-missing values after baseline and prior to the missing value.

For composite scores which consist of several component subscores, components will be imputed first and the composite score will be calculated based on imputed component subscores. Only if composite score still can't be calculated after imputing component subscores, then the previous composite score will be carried forward.

Missing Items on Questionnaire Scales:

Mayo scores (and alternative composite scores) can only be calculated if all subscores included are collected. If one or more subscores are missing, the score will be considered missing. No imputations will be made. The symptoms of stool frequency and rectal bleeding are scored as a 5 day average of daily values from up to 14 days preceding the visit, rounded to one decimal place. The most recent consecutive days will be used. If consecutive days are not available, then non-consecutive days can be used. Diary entries for stool frequency and rectal bleeding should not be included in the 5 days prior to the visit that are evaluated for these subscores for the following days: (1) the day the subject received medication for bowel preparation prior to endoscopy, (2) the day the subject underwent an endoscopy, and (3) 2 days following the endoscopy. Earlier diary entries will be used accordingly in order to provide the most recent data for 5 days prior to the respective study visit. If less than 3 days of diary data are available, stool frequency and rectal bleeding subscores will be considered missing.

When calculating **PUCAI**, if more than 3 subscores are missing then PUCAI will be considered missing. Otherwise, for missing subscores, the last available post-baseline value for the respective subscore will be carried forward to compute the total PUCAI score. The subscores of abdominal pain, rectal bleeding, stool consistency, number of stools per 24 hours, nocturnal stools and activity level are scored as a 2 day average, rounded to one decimal place. The most recent consecutive days will be used. If consecutive days are not available, then non-consecutive days can be used.

When calculating **IMPACT III** total score, if 4 or more of the items are missing then IMPACT III total score and all subscores will be considered missing. Otherwise, if there are 3 or fewer missing answers, the average of non-missing values will be used to replace the missing items for the total score and subscore calculation. If for subscores consisting of only 3 items (see 'Score Definitions' later in this Section), all of these items are missing and are replaced per above convention, a footnote will be added indicating for how many subjects the subscore included in the analysis has been fully imputed.

When calculating the **WPAI** scores, the following computational notes should be followed.

- Define Employment as a binary YES or NO variable where YES corresponds to "Employed" and NO corresponds to "Not Employed":
  - A subject will be considered EMPLOYED at a given visit if Q1 = YES or Q2 > 0 or Q4 > 0.
  - A subject will be considered UNEMPLOYED at a given visit if Q1 = NO and no positive hours recorded under Q2 and Q4 (i.e., if Q1 = NO AND Q2 ≤ 0 AND Q4 ≤ 0, then UNEMPLOYED).
  - Employment status for a subject will be considered missing at a given visit if Q1 = missing and no positive hours recorded under Q2 and Q4.
- If a subject is UNEMPLOYED or employment status is missing, then S1, S2, and S3 will be set to missing.
- If Q2 = 0 and Q4 = 0 or missing then Q2/(Q2 + Q4) = missing (i.e., S1 = missing).
- If Q2 = 0 and Q4 = 0, then set S3 to missing.
- If Q2 is missing or Q4 is missing, then set S1 and S3 to missing.
- If Q4 = missing, then DO NOT set Q5 = missing.
- If Q5 is missing, then apply the following rules:
  - If Q2 > 0, Q4 = 0, and Q5 = missing, then S3 = 100%.
  - If Q2 = 0, Q4 > 0, and Q5 = missing, then S3 is missing.
  - If Q2 > 0, Q4 > 0, and Q5 = missing, then S3 is missing.
- Determine if a subject missed work (based on Q2) in order to analyze the proportion of subjects who missed work:
  - Create a binary (YES or NO) "missed work" variable.
  - A subject will be considered as yes to missed work if Q2 is greater than 0.
  - If Q2 = missing, then MISSED WORK = missing.
  - If Q2 > 0, then MISSED WORK = YES
  - If Q2 = 0, then MISSED WORK = NO

- Therefore, the proportion of subjects who missed work will be counted based on the number of subjects with MISSED WORK = YES.

Partial Study Dates: If the day and/or month are missing, the following conventions will be used to impute the missing visit (or assessment) dates other than the dosing dates:

- 01 for missing start day,
- End of month for missing end day,
- January 1st will be used for a missing start month,
- December 31st for missing end month.

In case of partially missing AE start and stop dates, the dates will be imputed by comparing to first dose date of study medication so that the corresponding AEs will be made treatment-emergent whenever possible.

In case of missing or partially missing study drug dosing dates, the dates will not be imputed. Subjects will be considered as not receiving dose on that date.

Rounding

Rounding will be performed only for presentation of results. No rounding will be performed before or during analyses/calculations. The ROUND function of SAS will be used to round results for presentation.

The mean and median will be rounded for presentation to one decimal more than the data were entered into the database. For example, mean systolic blood pressure will be presented to one decimal place (125.2 mmHg) when it is recorded to integer level in the database (110 mmHg). The standard deviation will be rounded to two decimal places more than the data were entered into the database (e.g., 25.31 mmHg for systolic blood pressure). The minimum and maximum values will be presented as entered into the database.

Probabilities will be rounded to three decimal places before assignment of statistical significance and will be presented in rounded format. Probabilities that round to zero or are reported by SAS as zero will be presented as '< 0.001.'

Percentages will be rounded for presentation to one decimal place; e.g., the proportion 0.1244 will be reported in percent as 12.4%.

Score Definitions

1. Mayo Score and Component Scores

Table 1. Definition of Mayo Score and Component Scores

Term	Definition
Mayo Score	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3), rectal bleeding (0 – 3), physician's global assessment (0 – 3) and endoscopic subscore (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease.
Partial Mayo Score	Composite score of UC disease activity based on the subscores of stool frequency, rectal bleeding, and physician's global assessment and DOES NOT include the endoscopic subscore. This score ranges from 0 – 9 points with higher scores representing more severe disease.
9 point Mayo Score (without PGA)	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3), rectal bleeding (0 – 3), and endoscopic subscore (0 – 3). This score ranges from 0 – 9 points with higher scores representing more severe disease.
6 point Mayo Score (without PGA and endoscopy subscore)	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3) and rectal bleeding (0 – 3). This score ranges from 0 – 6 points with higher scores representing more severe disease.

Table 2. Definition of Endpoints Based on Mayo Score and Component Scores

Term	Definition
Clinical remission per Partial Mayo Score	PMS ≤ 2 and no individual subscore > 1
Clinical response per Partial Mayo Score	Decrease in PMS ≥ 2 points and ≥ 30% from Study M11-290-Baseline
Clinical remission per Mayo Score	Mayo score ≤ 2 and no individual subscore > 1
Clinical response per Mayo Score	Decrease in Mayo Score ≥ 3 points and ≥ 30% from Study M11-290-Baseline
Mucosal healing	Endoscopy subscore of 0 or 1
Clinical remission per 9 point Mayo Score (without PGA)	Defined as the 9 point Mayo Score (without PGA) ≤ 2 and no individual subscore > 1
Clinical remission per 6 point Mayo Score (without PGA and endoscopy subscore)	Defined as the 6 point Mayo Score (without PGA and endoscopy subscore) ≤ 1

2. PUCAI

Pediatric Ulcerative Colitis Activity Index (PUCAI) is calculated as the sum of six subscores [each being a 2 day average, see above] of PUCAI (abdominal pain [no pain = 0, pain can be ignored = 5, pain cannot be ignored = 10], rectal bleeding [none = 0, small amount with < 50% of stools = 10, small amount with most stools = 20, large amount with > 50% of stool content = 30], stool consistency of most stools [formed = 0, partially formed = 5, completely unformed = 10], number of stools per 24 hours [0 to 2 = 0, 3 to 5 = 5, 6 to 8 = 10, > 8 = 15], nocturnal stools [no = 0, yes = 10], activity level [no limitation = 0, occasionally limited = 5, severe restriction = 10]) with maximum total score 85. Higher scores represent more severe disease.

- PUCAI remission is defined as PUCAI < 10.
- PUCAI response is defined as a decrease in PUCAI ≥ 20 points from Study M11-290-Baseline.

3. IMPACT III

The IMPACT III questionnaire assesses quality of life in IBD related fields and comprises 35 closed questions with a 5-point Likert scale for all answers ('good'

through 'bad' answers scored by 5 through 1; range of total score 35 – 175, higher scores suggesting higher quality of life).

Total Score: sum of Q1 - Q35

Subscores:

- Bowel symptoms (7 items: Q1, Q3, Q10, Q19, Q21, Q25, Q31)
- Systemic symptoms (3 items: Q6, Q28, Q32)
- Emotional functioning (7 items: Q4, Q5, Q11, Q12, Q13, Q16, Q29)
- Social functioning (12 items: Q8, Q9, Q14, Q17, Q18, Q20, Q23, Q26, Q27, Q30, Q34, Q35)
- Body image (3 items: Q7, Q15, Q33)
- Treatment/interventions (3 items: Q2, Q22, Q24)

4. WPAI

The WPAI was developed to measure the effect of disease on productivity at work and its impact on daily activities. The questions ask about the impact of disease to the patient's caregiver within the past 7 days. Six questions asked on this questionnaire can be described briefly as follows:

Questions:

Q1. Currently employed (working for pay)? (Yes/No)

Q2. Hours missed from work due to PROBLEM.

Q3. Hours missed due to other reasons.

Q4. Hours actually worked.

Q5. Degree PROBLEM affected productivity while working.

Q6. Degree PROBLEM affected regular activities other than job.

WPAI scores are expressed as *percent impairment* based on the above questions. The four main impairment measures or scores are derived as follows:

Scores:

S1. ***Absenteeism***: Percent work time missed due to PROBLEM:

$$100 \times \left[\frac{Q2}{Q2 + Q4} \right]$$

S2. ***Presenteeism***: Percent impairment while working due to PROBLEM:

$$100 \times \left[\frac{Q5}{10} \right]$$

S3. Percent ***overall work impairment*** due to PROBLEM:

$$100 \times \left[\frac{Q2}{Q2+Q4} + \left\{ 1 - \frac{Q2}{Q2+Q4} \right\} \times \frac{Q5}{10} \right]$$

S4. Percent ***activity impairment*** due to PROBLEM: $100 \times \left[\frac{Q6}{10} \right]$

5. Other

Body Mass Index (BMI, in kg/m^2) will be obtained as:
BMI = weight (kg)/(height [m] × height [m])

"z" score for height velocity will be obtained as: (Observed height velocity [cm/yr] – mean height velocity for age and sex [cm/yr]/[SD of the mean]). NOTE: The mean and Standard Deviation (SD) will come from published table values: The CDC Growth Charts webpage of the Center for Disease Control and Prevention with derivation details on the "z" score: http://www.cdc.gov/growthcharts/percentile_data_files.htm, STATAGE.xls ('stature-for-age-chart,' for height) to be used.

Inappropriate Tanner stage will be defined depending on age and sex as follows: girls ≥ 14 years/boys ≥ 15 of age with a Tanner stage of 1; girls ≥ 16 years/boys ≥ 18 of age with a Tanner stage < 4; girls ≥ 19 years/boys ≥ 19 of age with a Tanner stage of < 5.1.

7.0 Demographics, Baseline Characteristics, Medical History, and Previous/Concomitant Medications

7.1 Demographic and Baseline Characteristics

Demographics and baseline disease characteristics and UC disease history values will be summarized using descriptive statistics. For the following characteristics marked by '#' the source data from Study M11-290 should be used for analyses (although some of the

items may also have been transferred from Study M11-290 into the Study M10-870 database).

Demographics and baseline characteristics:

- Subject demographics
  - Sex# (female, male)
  - Age [years] {using birth date#}
  - Age categories (< 13 years, ≥ 13 years)
  - Race# (White, Black, Asian, American Indian/Alaska Native, and Other)
  - Ethnicity# (Hispanic or Latino, Japanese, no ethnicity)
  - Geographic region# (North America, Western Europe, Eastern Europe)
  - Body Weight# [kg]
  - Body Weight categories (< 40, 40 – < 60, ≥ 60 kg)
  - Height# [cm]
  - BMI [kg/m^2]
  - Tobacco/Nicotine Use#* (current user, former user, never, or unknown)
  - Alcohol Use#* (current user, former user, never, or unknown)

* Data from Study M11-290-Baseline, to be clarified in a footnote to the table.

- PPD skin test/QuantiFERON®-TB Gold test/TB prophylaxis#
  - Tuberculin PPD skin test (negative, positive), positive defined as ≥ 5 mm
  - QuantiFERON-TB Gold test (negative, positive, indeterminate)
  - Combined result of Tuberculin PPD skin test and QuantiFERON-TB Gold test
  - Enrollment under TB prophylaxis (yes, no)
- Disease characteristics at Baseline
  - Duration of UC
    - In months: (Baseline date - date of diagnosis of UC# + 1)/(365.25/12)
    - In years: (Baseline date - date of diagnosis of UC# + 1)/365.25
  - Site of UC# (Left sided UC vs. Extensive UC/pancolitis)

- Mayo Score#
- Mayo Score at Study M11-290-Baseline#
- Partial Mayo Score# (PMS)
- PMS at Study M11-290-Baseline#
- Endoscopy Subscore#
- Rectal Bleeding Subscore#
- Physician's Global Assessment (PGA) Subscore#
- Stool Frequency Subscore#
- Number of stools in the last 24 hours prior to baseline# (i.e., last non-missing value $-1 \leq$ Rx Day ≤ 1)
- PUCAI#
- PUCAI at Study M11-290-Baseline#
- IMPACT III total score and subscores#
- WPAI# (Absenteeism, Presenteeism, Work Impairment, Activity Impairment)
- hs-CRP# [mg/L]

7.2 Medical History

Medical history data will be summarized and presented using System Organ Class (SOC) and Preferred Terms (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA) dictionary. The SOCs will be presented in alphabetical order and the PTs will be presented in alphabetical order within each SOC. The number and percentage of subjects with a particular PT will be summarized. Subjects reporting more than one PTs within a SOC will be counted only once for that SOC.

7.3 Prior Treatment and Concomitant Medications

Prior medications are those medications taken prior to the first dose of study drug in Study M11-290. This includes medications with a start date before the first study drug administration date in Study M11-290, regardless of the end date of these medications.

Medications taken on the day of the first dose of study drug in Study M11-290 are not counted as prior medications.

In cases where incomplete or missing medication dates are collected, a conservative approach will be taken such that if a medication could have been a prior medication, it will be counted as a prior medication.

Baseline medications are those medications that subjects were on at the time of the first dose of study drug in Study M10-870. This includes medications that started prior to the first dose of study drug in Study M10-870 and continued after the first dose of study drug or ended at the time of the first dose of study drug, as well as, medications that began on the day of the first dose of study drug in Study M10-870.

Concomitant medications are those medications, other than study drug, taken after the first dose of study drug in Study M10-870 and within 14 days of the last dose of study drug in Study M10-870. This includes medications with a start date between first study drug administration and last study drug administration + 14 days, as well as, medications with a start date prior to first dose of study drug and which are ongoing after first dose of study drug in Study M10-870. Medications taken on the day of the first dose of study drug in Study M10-870 are counted as concomitant medications.

In the situation where an incomplete or missing medication date is collected (start or end date), a conservative approach will be taken such that if a medication could have been a concomitant medication, it will be counted as a concomitant medication.

The medications will be coded by WHO DRUG dictionary. The number and percentage of subjects who had taken medications will be summarized by generic drug name for all medications as well as for those flagged as UC-related medications. No statistical tests will be performed.

The number and percentage of subjects with prior (prior to Study M11-290) anti-diarrheal therapy, NSAIDs, biologic DMARDs, synthetic DMARDs, IBD related antibiotics, anti-TNF biologics, non-anti-TNF biologics, systemic corticosteroids and among those the oral

corticosteroids, IMM and aminosalicylates will be summarized. In addition, four groups of baseline IMM and/or systemic corticosteroids use (IMM with systemic corticosteroids, IMM without systemic corticosteroids, systemic corticosteroids without IMM, and neither IMM nor systemic corticosteroids) will be tabulated.

8.0 Patient Disposition

Subject accountability overall and by current investigator will be summarized with number and percentage of subjects.

Subjects with any dose adjustments will be summarized as follows:

- Total number and percentage of subjects with EOW vs EW at Baseline
- Number and percentage of subjects with at least one dose escalation (EOW to EW)
- Number and percentage of subjects with EW dosing (i.e., those who started on EW or who escalated to EW) with at least one de-escalation (EW to EOW)
- Number and percentage of subjects who de-escalated from EW to EOW with at least one re-escalation afterwards (EOW to EW)

Subject final status including the following categories will be summarized by number and percentage:

- Subjects who completed the study (defined as: either completed Week 288 or switched to adult care or to commercial Humira), also by sub-category of completion:
  - Completed Week 288
  - Switched to adult care
  - Switched to commercial Humira
- Subjects who discontinued

In addition, the reasons for premature discontinuation of study drug (primary reason and all reasons) will be summarized with frequencies and percentages. Subjects may have

multiple reasons for prematurely discontinuing study drug, but will be counted only once for the primary reason.

9.0 Study Drug Exposure and Compliance

The duration (days) of study drug exposure and treatment compliance (%) will be summarized using descriptive statistics.

Study drug exposure will be calculated as follows:

> ADA exposure in Study M10-870 = last dose date of ADA in Study M10-870 – first dose date of ADA in Study M10-870 + 14 days
>
> Overall ADA exposure in Study M11-290 & Study M10-870 = ADA exposure in Study M10-870 + Overall ADA exposure in Study M11-290 (see Study M11-290 SAP)

Treatment compliance (%) is defined as the number of study drug injections received divided by the number of injections planned.

10.0 Efficacy Analysis

10.1 General Considerations

No confirmatory testing for efficacy endpoints will be performed in this open-label study.

Endpoints that are of the binary type will be analyzed as proportions including 95% CIs. Endpoints that are of the continuous type will be analyzed as changes from baseline, and reported including 95% CIs. OC and LOCF (see Section 6.0) will be used, unless otherwise noted.

10.2 Efficacy Analysis

The efficacy endpoints are:

- The proportion of subjects who achieve clinical remission as measured by PMS (defined as a PMS ≤ 2 and no individual subscore > 1) over time;
- The proportion of subjects who achieve clinical response as measured by PMS (defined as a decrease in PMS ≥ 2 points and ≥ 30% from Study M11-290-Baseline) over time;
- The proportion of subjects who achieve PUCAI remission (defined as < 10) over time;
- The proportion of subjects who achieve PUCAI response (defined as a decrease in PUCAI ≥ 20 points from Study M11-290-Baseline).
- Change from Baseline in total IMPACT III Quality of Life score and subscores over time for subjects at least 9 years old at Baseline;
- Change from Baseline in WPAI scores over time;
- Change from Baseline in "z" scores for height over time;
- Change from Baseline in BMI over time;
- Change from Baseline in "z" scores for weight-for-age over time;
- Change from Baseline in albumin and total protein over time;
- Change from Baseline in hemoglobin, hematocrit, red blood cell count over time;
- Change from Baseline in hs-CRP levels over time;
- Proportion of subjects at appropriate Tanner stage (see Section 6.0) over time;
- Proportion of subjects in Mayo score clinical response over time (for subjects with available Mayo score);
- Proportion of subjects in Mayo score clinical remission over time (for subjects with available Mayo score);
- Proportion of subjects who achieve mucosal healing over time as measured by Mayo endoscopy subscore (defined as ≤ 1, for subjects with available Mayo score)
- Proportion of subjects in 9 point Mayo score (without PGA) clinical remission over time (for subjects with available Mayo score);

- Proportion of subjects in 6 point Mayo score (without PGA and endoscopy subscore) clinical remission over time (for subjects with available Mayo score).

10.3 Handling of Multiplicity

Not applicable since no confirmatory testing is performed.

10.4 Efficacy Sensitivity Analyses

Sensitivity analyses may be performed as deemed appropriate.

10.5 Efficacy Subgroup Analysis

The subgroups listed below will be used for proportions of subjects with PMS remission and Mayo score clinical remission over time.

- Sex (male, female)
- Age (< 13 years, ≥ 13 years)
- Geographic region (North America, Western Europe, Eastern Europe)
- Prior (prior to Study M11-290) anti-TNF use (yes, no)
- Baseline systemic corticosteroid use (yes, no)
- Baseline IMM use (yes, no)
- Weight (< 40 kg, ≥ 40 kg)
- Pancolitis (yes, no)
- Disease duration (≤ Baseline-median, > Baseline-median)

11.0 Safety Analysis

11.1 General Considerations

Incidence of adverse events (AEs), changes in vital signs, physical examination results, and clinical laboratory data are the safety parameters in this study.

11.2 Analysis of Adverse Events

Treatment emergent AEs, serious adverse events (SAEs), and AEs of special interest will be summarized.

AEs of special interest include:

- Infections
- Serious infections
- Opportunistic infections excluding oral candidiasis and tuberculosis (TB)
- TB (active/latent)
- Oral candidiasis
- Legionella infections
- Parasitic infections
- Diverticulitis
- Malignancies
- Lymphoma
- Hepatosplenic T-cell lymphoma (HSTCL)
- Leukemia
- Non-melanoma skin cancer (NMSC)
- Melanoma
- Malignancy other than lymphoma, HSTCL, leukemia, NMSC or melanoma
- Lupus-like reactions and systemic lupus erythematosus
- Allergic reactions including angioedema/anaphylaxis
- Stevens-Johnson syndrome
- Sarcoidosis
- Vasculitis (cutaneous/non-cutaneous)
- Demyelinating disorder
- Interstitial lung disease
- Congestive heart failure (CHF)
- Myocardial infarction (MI)

- Cerebrovascular accident (CVA)
- Intestinal perforation
- Pancreatitis
- Hematologic disorders including pancytopenia
- Liver failure and other liver events
- Reactivation of Hepatitis B
- Autoimmune Hepatitis
- Injection site reaction
- Erythema multiforme
- Worsening or new onset of psoriasis
- Pulmonary embolism
- Progressive multifocal leukoencephalopathy (PML)
- Reversible posterior leukoencephalopathy syndrome (RPLS)
- Amyotrophic lateral sclerosis
- Humira administration related medication error

Treatment-emergent AEs are defined as events with an onset date after the first dose of the study drug in Study M10-870 (i.e., on/after Rx Day 1) and with an onset date no more than 70 days after the last dose of the study drug in Study M10-870.

AEs with missing or unknown severity will be categorized as severe. AEs with missing or unknown relationship to study drug will be categorized as 'Reasonable Possibility' of being related to study drug. AEs that are reported more than 70 days after last dose of study drug will be excluded from the AE summaries; however, all reported AEs will be included in the AE data listings.

Overview tables of number and percentage of subjects with treatment-emergent AEs and events per 100 patient-years of ADA exposure in Study M10-870 (as defined in Section 9.0) will be provided.

AE data will be summarized and presented using primary MedDRA system organ classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used to code the AE data. The system organ classes will be presented in alphabetical order and the preferred terms will be presented in alphabetical order within each system organ class.

In addition, a summary of AEs by maximum severity and relationship to study drug will be presented.

A subject who reports more than one AE in different system organ classes will be counted only once in the overall total. A subject who reports two or more different preferred terms which are in the same SOC will be counted only once in the SOC total. A subject who reports more than one AE with the same preferred term will be counted only once for that preferred term using the most extreme incident (i.e., most "severe" for the severity tables and as having a 'Reasonable Possibility' of being related to study drug for the relationship tables).

Summaries by subject number and listings will also be provided as appropriate.

An overview of AE categories by type of output is provided below:

- Adverse Event Overview

  The number and percentage of subjects experiencing treatment-emergent AEs, as well as events per 100 PYs, will be summarized for the following AE categories:
  - Any AE
  - SAEs
  - Severe AEs
  - AEs leading to discontinuation of study drug
  - AEs rated as possibly related to study drug by the investigator (reasonable possibility)

- SAEs rated as possibly related to study drug by the investigator (reasonable possibility)
- AEs leading to Death
- AEs of special interest

- Adverse Events by System Organ Class and Preferred Term

  The number and percentage of subjects experiencing treatment-emergent AEs, as well as events per 100 PYs, will be summarized by SOC and PT for the following AE categories.
  - Any AE
  - SAEs
  - AEs leading to discontinuation of study drug
- Subject Numbers Associated with treatment-emergent Adverse Events by System Organ Class and Preferred Term
  - Any AE
- Adverse Event listing
  - Any AE
  - SAEs
  - Severe AEs
  - AEs leading to discontinuation of study drug
  - AEs rated as possibly related to study drug by the investigator (reasonable possibility)
  - SAEs rated as possibly related to study drug by the investigator (reasonable possibility)
  - AEs leading to Death
  - AEs of special interest

11.3 Analysis of Laboratory Data

Hematology	Clinical Chemistry	Urinalysis[a]
Hematocrit Hemoglobin Red Blood Cell (RBC) count White Blood Cell (WBC) count Neutrophils Bands Lymphocytes Monocytes Basophils Eosinophils Platelet count (estimate not acceptable)	Blood Urea Nitrogen (BUN) Creatinine Total bilirubin Albumin Serum glutamic-pyruvic transaminase (SGPT/ALT) Serum glutamic-oxaloacetic transaminase (SGOT/AST) Alkaline phosphatase Sodium Potassium Calcium Inorganic phosphorus Uric acid Cholesterol Total protein Glucose Triglycerides	Specific gravity Ketones pH Protein Glucose Blood **Others** High-sensitivity C-reactive protein (hs-CRP) β-HCG

For selected continuous clinical laboratory parameters, change from baseline to minimum (smallest) value, maximum (largest) value, and final value during ADA exposure in Study M10-870 will be summarized using descriptive statistics.

A listing of all subjects with any clinical laboratory determinations during the study meeting criteria for potentially clinically importance (Common Toxicity Criteria [CTC] Grade ≥ 3, based on the National Cancer Institute Common Toxicity Criteria for Adverse Event (NCI CTCAE) scale, most recent version) will be provided. For each of these subjects, the whole course of the parameter will be listed.

CTC Grading derived from NCI CTCAE v. 5.0

Test	Grade 1	Grade 2	Grade 3	Grade 4
Hemoglobin decreased	< LLN - 100 g/L	< 100 – 80 g/L	< 80 g/L	-
White blood cell decreased	< LLN - 3.0 × 10^9/L	< 3.0 - 2.0 × 10^9/L	< 2.0 - 1.0 × 10^9/L	< 1.0 × 10^9/L
Neutrophil count decreased	< LLN - 1.5 × 10^9/L	< 1.5 - 1.0 × 10^9/L	< 1.0 - 0.5 × 10^9/L	< 0.5 × 10^9/L
Lymphocyte count decreased	< LLN - 0.8 × 10^9/L	< 0.8 - 0.5 × 10^9/L	< 0.5 - 0.2 × 10^9/L	< 0.2 × 10^9/L
Platelet count decreased	< LLN - 75.0 × 10^9/L	< 75.0 - 50.0 × 10^9/L	< 50.0 - 25.0 × 10^9/L	< 25.0 × 10^9/L
Creatinine increased	> ULN - 1.5 × ULN	> 1.5 - 3.0 × baseline; > 1.5 - 3.0 × ULN	> 3.0 × baseline; > 3.0 - 6.0 × ULN	> 6.0 × ULN
Bilirubin increased	> ULN - 1.5 × ULN if baseline was normal; > 1.0 - 1.5 × baseline if baseline was abnormal	> 1.5 - 3.0 × ULN if baseline was normal; > 1.5 - 3.0 × baseline if baseline was abnormal	> 3.0 - 10.0 × ULN if baseline was normal; > 3.0 - 10.0 × baseline if baseline was abnormal	> 10.0 × ULN if baseline was normal; > 10.0 × baseline if baseline was abnormal
SGPT/ALT increased	> ULN - 3.0 × ULN if baseline was normal; 1.5 - 3.0 × baseline if baseline was abnormal	> 3.0 - 5.0 × ULN if baseline was normal; > 3.0 - 5.0 × baseline if baseline was abnormal	> 5.0 - 20.0 × ULN if baseline was normal; > 5.0 - 20.0 × baseline if baseline was abnormal	> 20.0 × ULN if baseline was normal; > 20.0 × baseline if baseline was abnormal
SGOT/AST increased	> ULN - 3.0 × ULN if baseline was normal; 1.5 - 3.0 × baseline if baseline was abnormal	> 3.0 - 5.0 × ULN if baseline was normal; > 3.0 - 5.0 × baseline if baseline was abnormal	> 5.0 - 20.0 × ULN if baseline was normal; > 5.0 - 20.0 × baseline if baseline was abnormal	> 20.0 × ULN if baseline was normal; > 20.0 × baseline if baseline was abnormal
Alkaline phosphatase increased	> ULN - 2.5 × ULN if baseline was normal; 2.0 - 2.5 × baseline if baseline was abnormal	> 2.5 - 5.0 × ULN if baseline was normal; > 2.5 - 5.0 × baseline if baseline was abnormal	> 5.0 - 20.0 × ULN if baseline was normal; > 5.0 - 20.0 × baseline if baseline was abnormal	> 20.0 × ULN if baseline was normal; > 20.0 × baseline if baseline was abnormal

Shift tables from baseline to worst CTC grade will be provided.

The liver-specific laboratory tests include the serum glutamic pyruvic transaminase (SGPT/ALT), serum glutamic-oxaloacetic transaminase (SGOT/AST), alkaline phosphatase, and total bilirubin. A listing of potentially clinically significant liver function laboratory values will be provided. The listing will include all subjects who met any of the following 4 criteria:

- ALT ≥ 3.0 × ULN, or
- AST ≥ 3.0 × ULN, or
- Alkaline phosphatase ≥ 1.5 × ULN, or
- Total bilirubin ≥ 2.0 × ULN.

11.4 Analysis of Vital Signs and Weight

The following vital sign parameters will be obtained at each visit and summarized:

- Sitting systolic blood pressure (mmHg)
- Sitting diastolic blood pressure (mmHg)
- Sitting heart rate (or pulse) (bpm)
- Weight (kg)
- Height (cm)
- Body Mass Index (kg/m^2)
- Respiratory rate (rpm)
- Body temperature (°C)

For continuous vital sign parameters, mean change from Baseline to minimum (smallest) value, maximum (largest) value, and final value during ADA exposure in Study M10-870 will be summarized. Subjects with potentially clinically significant results during ADA exposure in Study M10-870 will be identified according to the criteria in Table 3. Vital sign results meeting the criteria for potentially clinically significant findings will also be identified in a listing.

Table 3. Criteria for Potentially Clinically Significant Vital Sign Results

Vital Sign	Criterion	Definition of Potentially Clinically Significant
Systolic blood pressure	Low	Value ≤ 70 mmHg and decreased ≥ 20 mmHg from initial value
	High	Value ≥ 160 mmHg and increased ≥ 20 mmHg from initial value
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decreased ≥ 15 mmHg from initial value
	High	Value ≥ 105 mmHg and increased ≥ 15 mmHg from initial value
Pulse	Low	Value ≤ 50 bpm and decreased ≥ 30 bpm from initial value
	High	Value ≥ 120 bpm and increased ≥ 30 bpm from initial value

12.0 Summary of Changes

Not applicable.

13.0 References

1. Klein DA, Emerick JE, Sylvester JE, et al. Disorders of Puberty: An Approach to Diagnosis and Management. Am Fam Physician. 2017;96(9):590-9.

Abbreviations

ADA	Adalimumab
AE	Adverse event
BMI	Body mass index
CI	Confidence Interval
CTC	Common Toxicity Criteria
DB	Double-blinded
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
ECG	Electrocardiogram
EOW	every other week
EW	every week
ESR	Erythrocyte sedimentation rate
HACA	Human anti-chimeric antibody
hs-CRP	High Sensitivity C-Reactive protein
ICH	International Conference on Harmonization
IDMC	Independent Data Monitoring Committee
ITT	Intent-to-treat
LOCF	Last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified intent-to-treat
MTX	Methotrexate
NRI	Non-Responder imputation
OC	Observed case
PGA	Physician's global assessment
PK	Pharmacokinetics
PMS	Partial mayo score
PUCAI	Pediatric Ulcerative Colitis Activity Index
RR	Re-Randomized
SAE	Serious adverse event
SAP	Statistical analysis plan
SC	Subcutaneous
TB	Tuberculosis
TNF	Tumor necrosis factor

UC	Ulcerative Colitis