1.0 TITLE PAGE

[redacted]

MT10109L-001

A Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel-group Study to Evaluate the Safety and Efficacy of MT10109L (NivobotulinumtoxinA) for the Treatment of Glabellar Lines

STATISTICAL ANALYSIS PLAN - Clinical Study Report

[Final]: [18 NOV 2020]

2.0 TABLE OF CONTENTS

1.0 TITLE PAGE 1

2.0 TABLE OF CONTENTS 2

3.0 LIST OF ABBREVIATIONS 4

4.0 INTRODUCTION 6

5.0 OBJECTIVES 11

6.0 PARTICIPANT POPULATIONS 12
6.1 Intent-to-Treat Population 12
[redacted]
6.3 Safety Population 12

7.0 PARTICIPANT DISPOSITION 13

8.0 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS 14

9.0 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE 15
9.1 Extent of Exposure 15

10.0 EFFICACY ANALYSES 16
10.1 Primary Efficacy Parameters 16
10.2 Secondary Efficacy Parameters 18
10.3 Exploratory Efficacy Parameters 20

11.0 SAFETY ANALYSES 22
11.1 Adverse Events 22
11.2 Clinical Laboratory Evaluations 24
11.3 Vital Signs 25
11.4 Electrocardiogram 25
[redacted]
11.7 Pregnancy Tests 26
[redacted]

12.0 INTERIM ANALYSIS 27

13.0 DETERMINATION OF SAMPLE SIZE 28

14.0 STATISTICAL SOFTWARE 29

15.0 DATA HANDLING CONVENTIONS 30
[redacted]
15.2 Multiple Imputation Method 33
[redacted]
15.5 Repeated or Unscheduled Assessments of Safety Parameters 36
15.6 Missing Severity Assessment for Adverse Events 37
15.7 Missing Causal Relationship to Study Intervention for Adverse Events 37

15.8 Missing Date Information for Adverse Events....37
15.9 Missing Date Information for Prior or Concomitant Medications....37

[redacted]

17.0 REFERENCES....39

18.0 HISTORY OF CHANGES....40

3.0 LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse events of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
CMH	Cochran-Mantel-Haenszel
CRF	case report form
ECG	electrocardiogram, electrocardiographic
EU	European Union
eCRF	electronic case report form
FLO-11	Facial Line Outcomes Questionnaire
FLSQ	Facial Line Satisfaction Questionnaire
FWS	Facial Wrinkle Scale
GL	glabellar lines
ITT	intent to treat
LCL	lateral canthal lines
[redacted]	[redacted]
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent-to-treat
[redacted]	[redacted]
Q1	25th percentile
Q3	75th percentile
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
RBC	Red blood cell

SAE	serious adverse event
SAP	statistical analysis plan
SAS	Statistical Analysis Software
SI	*Le Système International d'Unités* (International System of Units)
TEAE	treatment-emergent adverse event
WBC	white blood cell

4.0 INTRODUCTION

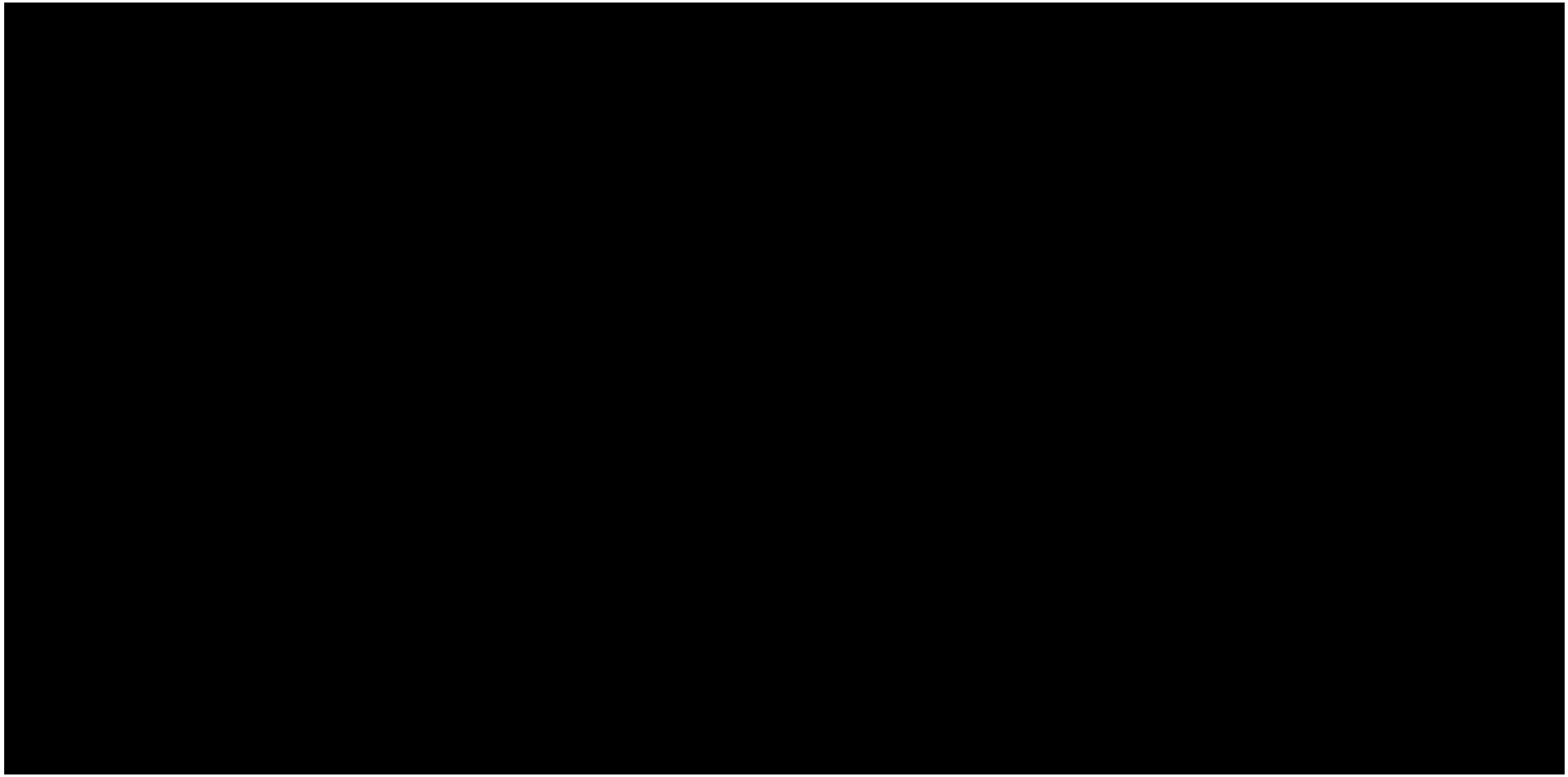
This statistical analysis plan (SAP) provides a more technical and detailed elaboration of the statistical analyses of the efficacy and safety data as outlined and/or specified in the final protocol amendment of Study MT10109L-001 (dated Oct 2020). Specifications of tables, figures, and data listings are contained in a separate document.

Study MT10109L-001 is a multicenter, randomized, double-blind, placebo-controlled, parallel-group Phase 3 study conducted across approximately 16 global sites to evaluate the safety and efficacy of MT10109L in treating glabellar lines (GL). Participants may receive up to 3 study interventions (a single double-blind intervention of MT10109L 20 U or placebo and up to 2 open-label interventions with MT10109L 20 U). Participants are adults ≥ 18 years of age with moderate to severe GL at maximum frown (assessed by both the investigator and participant using the FWS [investigator and participant ratings must be the same for GL]).

The total duration of study participation for each participant is approximately 12 months. On Day 1, participants will be randomly assigned in a 2:1 ratio to receive MT10109L 20 U or placebo. Enrollment stratification will be by baseline GL severity at maximum frown, assessed by the clinician (investigator or subinvestigator) using the FWS.

[redacted]

[redacted]



5.0 OBJECTIVES

The objectives of this study are to evaluate the safety and efficacy between 20 U MT10109L and placebo for the treatment of participants with moderate to severe glabellar lines (GL).

6.0 PARTICIPANT POPULATIONS

6.1 INTENT-TO-TREAT POPULATION

The Intent-to-Treat (ITT) population consists of all randomized participants. Primary and secondary efficacy analyses for the US FDA will be performed on the ITT population.

[redacted]

6.3 SAFETY POPULATION

The safety population will include all participants who receive at least 1 injection of study intervention. Safety analyses will be based on safety population. All safety analyses will be performed with participants by their actual intervention received.

7.0 PARTICIPANT DISPOSITION

The number and percentage of participants in all 3 analysis populations (ITT, [redacted] safety) will be summarized by treatment group and overall; the number of participants screened and randomized will be summarized by study center.

The number of participants screened will be summarized overall. The number and percentage of participants who were randomized, treated, or who completed the study and prematurely discontinued will be presented for each treatment group [redacted]. A frequency table showing participant disposition [redacted] will also be provided for all analysis populations for each treatment cycle.

Tabulation of the numbers and percentages of participants in each exit status category (i.e., adverse event, pregnancy, lost to follow-up, personal reasons, protocol violations and other) will be provided for each treatment group for entire study and by treatment cycle for all analysis populations. Discontinued participants will be listed along with the corresponding reason(s) for early withdrawal from the study.

8.0 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographic parameters and baseline characteristics will be summarized by treatment group for ITT [redacted]. Continuous variables will be summarized by number of participants and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of participants.

Demographic parameters include age, age group (< 65 and ≥ 65), race, ethnicity and sex.

Baseline characteristics include weight, height, body mass index (BMI, calculated as weight [kg]/ height[m]^2), FWS scores of GL severity at maximum frown and at rest (assessed by investigators and participants), [redacted]. The distribution of the baseline characteristics will be summarized by treatment group.

The number and percentage of participants with abnormalities in medical and surgical histories in each system organ class (SOC) and preferred term (PT) will be summarized by treatment group for the ITT population.

Prior medications include all medications taken any time prior to the Day 1 baseline visit, whether or not medication is continuing beyond the Baseline Visit.

Concomitant medications encompass all medicinal products that the participant was taking prior to the Day 1 baseline visit which are ongoing at the visit, in addition to all medications that have a start date on or after the Day 1 visit date.

WHODrug [redacted] will be used to classify prior and concomitant medications by therapeutic class and drug name.

Both prior and concomitant medications will be coded by drug name and therapeutic class. The use of prior and concomitant medications will be summarized by drug class and drug name in each treatment group for the ITT population.

9.0 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

9.1 EXTENT OF EXPOSURE

Participants' exposure to study intervention will be summarized by total duration of intervention exposure and by cycle [redacted]. In addition, for the entire study, the number and percentage of participants receiving [redacted] study interventions will be summarized.

10.0 EFFICACY ANALYSES

For US FDA, the primary and secondary efficacy analyses will be based on the ITT population. For EU regulatory agencies, the primary and secondary efficacy analyses will be based on the mITT population. Exploratory efficacy analyses will be based on the ITT population.

Missing values for the primary measures will be imputed using multiple imputation methods up to Day 180 for Treatment Cycle 1. [redacted]

The evaluation of the equality of the proportions of responders will be based on Cochran-Mantel-Haenszel (CMH) test stratified by investigator-assessed baseline GL severity at maximum frown. [redacted]

Continuous descriptive statistics include: N1 (number of participants with non-missing values at both baseline and the specified post-baseline analysis visit), mean, standard deviation (SD), median, minimum, and maximum.

Categorical variables will be summarized by number and percentage of participants.

Efficacy analyses will be done for all visits. The primary timepoint is Day 30 of Treatment Cycle 1.

The 95% confidence intervals for the treatment difference in response rates and relative risk will be summarized [redacted]. The p-value for CMH test will be presented as well. [redacted]

10.1 PRIMARY EFFICACY PARAMETERS

US FDA:

The US FDA composite primary efficacy endpoint is the proportion of participants with a $\geq$ 2-grade improvement from baseline on the FWS according to both the investigator and participant-rated assessments of GL severity at maximum frown at Day 30 of Treatment Cycle 1.

The primary analysis will be performed on the ITT population.

The following hypothesis will be used to compare the MT10109L groups with placebo:

- Null hypothesis: MT10109L 20 U and placebo are equally effective in reducing GL severity at maximum frown as measured by the proportion of responders with a ≥ 2-grade improvement from baseline based on both the investigator- and participant-rated FWS at Day 30 of Treatment Cycle 1.

- Alternative hypothesis: MT10109L 20 U and placebo are not equally effective in reducing GL severity at maximum frown as measured by the proportion of responders with a ≥ 2-grade improvement from baseline based on both the investigator- and participant-rated FWS at Day 30 of Treatment Cycle 1.

EU Regulatory agencies:

[redacted]

[redacted]

10.2 SECONDARY EFFICACY PARAMETERS

US FDA:

For US FDA, the hierarchical ranking of the secondary efficacy endpoints is as follows:

1. The duration of GL treatment effect estimated as the median time to return to *moderate* or *severe* GL at maximum frown, in participants who achieved a rating of ≥ 2-grade improvement from baseline in GL severity at maximum frown at Day 30 Treatment Cycle 1 according to investigator assessments using the FWS (this will not be included in the hierarchical testing strategy for multiplicity control)

2. The proportion of responders for investigator assessments of GL severity at maximum frown using the FWS, where a responder is defined as achieving a ≥ 2-grade improvement from baseline at maximum frown at Day 30 of Treatment Cycle 1

3. [redacted]

4. The proportion of participants reporting *mostly satisfied/very satisfied* on the FLSQ follow-up version Item 5 for GL at Day 60 of Treatment Cycle 1

5. The proportion of responders for investigator assessments of GL severity at rest using the FWS, among participants who were rated at least *mild* at rest at baseline, where a responder is defined as achieving a ≥ 1-grade improvement from baseline at Day 30 of Treatment Cycle 1

EU Regulatory Agencies

[redacted]

Analyses of the secondary efficacy variables will be performed for all study visits using observed data, with the primary time-point at Day 30 of Treatment Cycle 1 [redacted].

The duration of GL treatment effect is defined as the time to return to *moderate* or *severe* GL at maximum frown in those participants achieving a responder status (for US FDA, responder is defined as participants who achieved a rating of ≥ 2-grade improvement from baseline in GL severity at maximum frown; [redacted] It will be estimated using the Kaplan-Meier method. [redacted]

[redacted]

A serial gatekeeping strategy will be used to account for multiplicity in testing the primary and secondary efficacy variables at the primary timepoint [redacted]

The proportion of responders will be analyzed using the CMH test stratified by Baseline GL severity at maximum frown assessed by investigator [redacted]

10.3 EXPLORATORY EFFICACY PARAMETERS

For both the US FDA and EU, analyses for exploratory efficacy endpoints will be performed on the ITT population. The proportion of responders will be analyzed using the CMH test [redacted]

[redacted]

11.0 SAFETY ANALYSES

The safety analysis will be performed using the safety population. The safety parameters will include adverse events (AEs), [redacted] vital signs, Electrocardiogram (ECG), and immunogenicity analyses. For each safety parameter of the vital sign and ECG parameters, the last non-missing safety assessment before the first dose of study intervention will be used as the baseline for all analyses of that safety parameter. Continuous variables will be summarized by number of participants and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of participants.

11.1 ADVERSE EVENTS

AEs will be coded from the verbatim text into preferred term (PT) and the primary system organ class (SOC) by using the MedDRA dictionary. In general, adverse events (AEs) data will be analyzed and presented for:

1) TEAEs: An adverse event will be considered a treatment-emergent adverse event (TEAE) if: 1) The adverse event began on or after the date of the first study intervention; or 2) The adverse event was present before the date of the first study intervention, but increased in severity or became serious on or after the date of the first study intervention. An adverse event that occurs more than 30 days after the study exit will not be counted as a TEAE.

[redacted]

Per case report form instructions, a new AE record will be created with a new AE onset date for any AE that worsens. Therefore, TEAEs can simply be identified as those AEs with recorded onset date on or after the date of the first study intervention and within 30 days after the study exit.

In each of the analysis periods (entire study, or by MT10109L cycle), a specific TEAE will only count once per participant, associated with its worst severity during the time period of interest. Unless stated otherwise, the method of analyses described in this section will be applied to each of the screening/baseline and study treatment periods.

Adverse events will be summarized by treatment group for entire study and by MT10109L cycle in descending order of incidence rate. [redacted]

[redacted]

Three incidence rate tables are presented for summarizing all TEAEs:

1) by descending order of incidence rate
2) by primary SOC and PT
3) by primary SOC, PT, and severity

Serious TEAEs and treatment-related TEAEs will be summarized by primary SOC and PT for the entire study and by MT10109L cycle. Treatment-related TEAEs will also be summarized by related to study intervention or related to study procedure for the entire study and by MT10109L cycle. TEAEs leading to study discontinuation will be summarized by primary SOC and PT for entire study and by MT10109L cycle. [redacted]

A participant listing will be generated for all AEs, SAEs, treatment-related AEs and AEs leading to study discontinuation.

[redacted]

[redacted]

[redacted]

11.2 CLINICAL LABORATORY EVALUATIONS

Hematology and non-fasting blood chemistry assays will be collected at screening visit only and performed by a central laboratory. The clinical laboratory parameters include the following:

Hematology:	Hemoglobin, hematocrit, red blood cell count (RBC), RBC morphology, white blood cell count (WBC), neutrophils, lymphocytes, monocytes, basophils, eosonophils, and platelets

Blood Chemistry: glucose, creatinine, urea nitrogen, total bilirubin, albumin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase, uric acid, sodium, potassium, bicarbonate (carbon dioxide content), chloride, phosphorus, calcium, magnesium, and total protein.

A participant level listing for hematology and blood chemistry will be provided.

11.3 VITAL SIGNS

Study baseline is defined as the data measured before dosing on Day 1. Baseline and change from study baseline data on blood pressure (mm Hg), pulse rate (beats/min) and respirations (breaths per minute) will be summarized as descriptive statistics for each time point and each visit for each treatment group.

11.4 ELECTROCARDIOGRAM

Descriptive statistics for ECG parameters (eg, heart rate, PR interval, RR interval, QRS duration, QT interval, corrected QT [QTc] intervals) at baseline, and changes from baseline at all post-baseline timepoints will be presented for all participants in the Safety Population. For each parameter, only participants who had both baseline post-baseline assessments will be included in the summary. [redacted]

[redacted]

[redacted]

11.7 PREGNANCY TESTS

Urine pregnancy tests are performed prior to treatment and at exit visit for females of childbearing potential. Participants with positive pregnancy test results will be listed by treatment group, including urine sample collection date, and days since most recent treatment.

[redacted]

[redacted]

[redacted]

[redacted]

12.0 INTERIM ANALYSIS

No interim analysis is planned for this study.

13.0 DETERMINATION OF SAMPLE SIZE

Approximately 225 participants will be randomized into the study in a 2:1 ratio [redacted] The sample size of 225 allows for an adequate safety database of participants treated with MT10109L for this indication. Based on previous clinical studies with BOTOX for the treatment of facial lines, the dropout rate is estimated to be 15% [redacted]

14.0 STATISTICAL SOFTWARE

Statistical analyses will be performed using version 9.3 (or newer) of SAS [redacted]

15.0 DATA HANDLING CONVENTIONS

- The level of significance used for all statistical tests will be 0.05, 2-sided, unless stated otherwise. P-values ≤ 0.050 will be considered as statistically significant.

[redacted]

- Descriptive statistics for continuous variables include the sample size (N), mean, standard deviation, median, minimum (min), and maximum (max).

- Summary statistics for categorical variables include the sample size (N), frequency count, and percent.

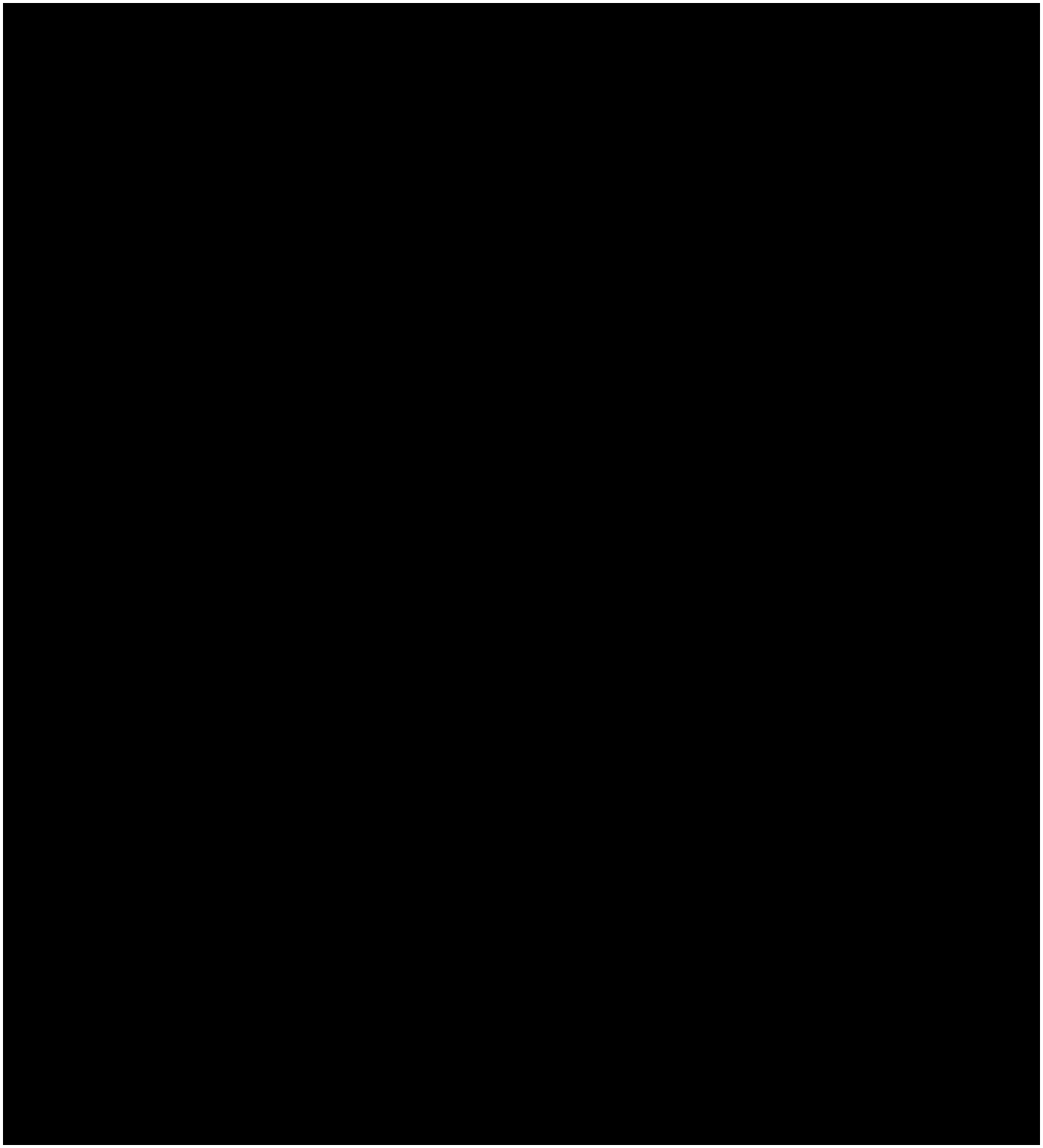
[redacted]

- The Medical Dictionary for Regulatory Activities (MedDRA) will be used to code all adverse events (AEs) and medical history.

- World Health Organization Drug Dictionary Enhanced (WHO DDE) preferred name and MedDRA will be used to code all medications.

[redacted]

[redacted]

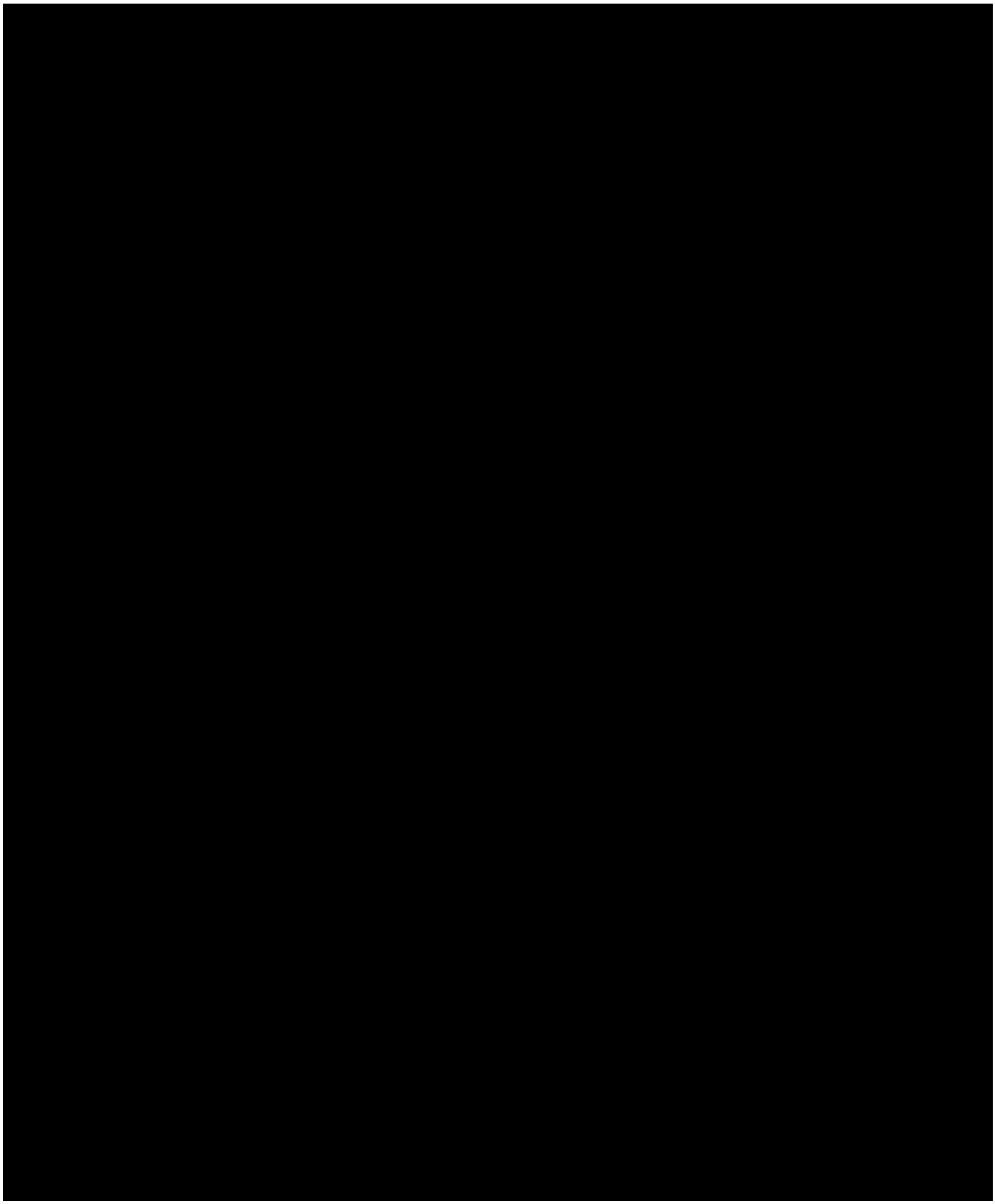


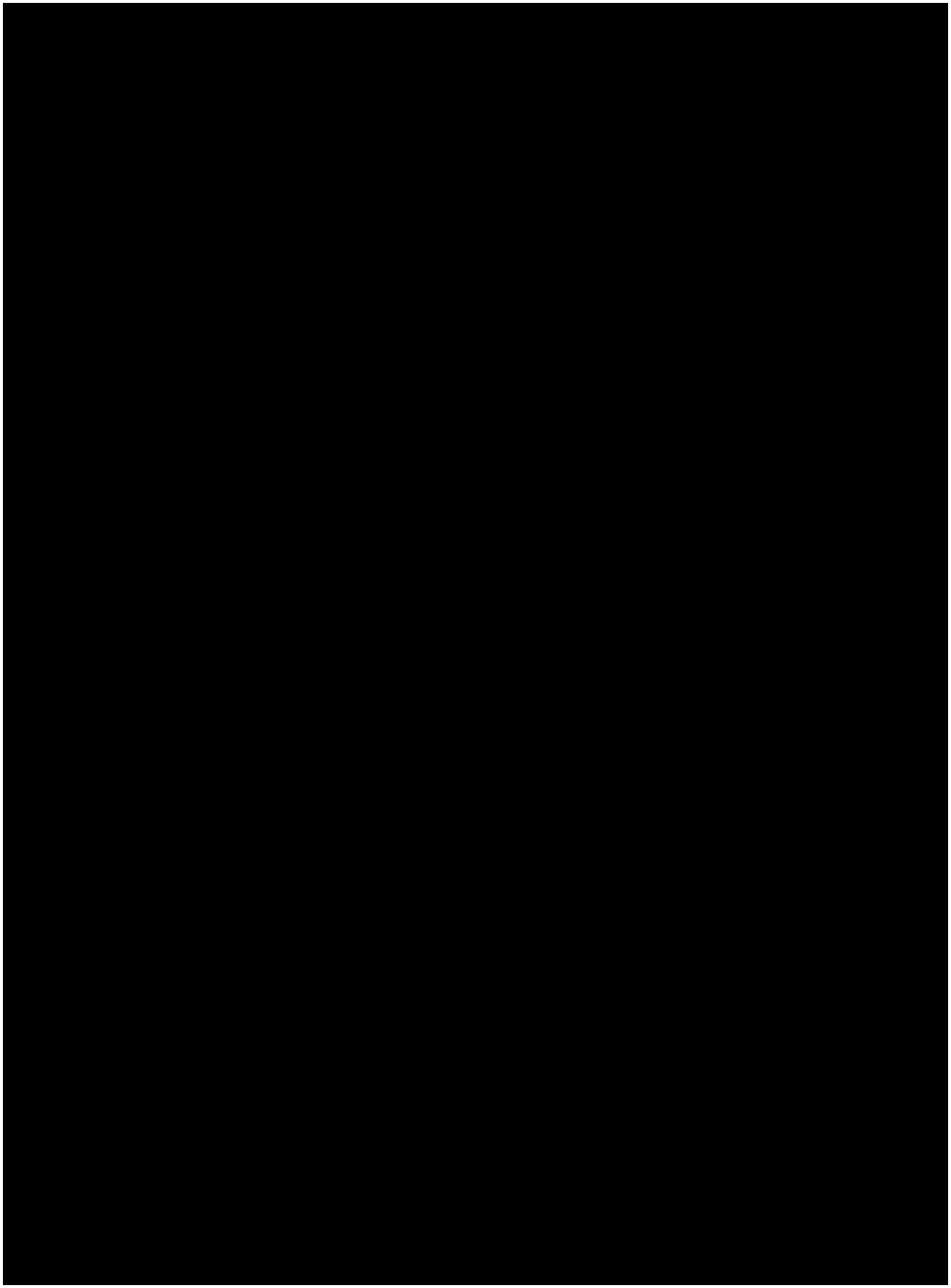
[redacted]

15.2 MULTIPLE IMPUTATION METHOD

SAS Proc MI procedure will be used [redacted]. Seed for all Proc MI procedure is pre-specified [redacted]

[redacted]





[redacted]

15.5 REPEATED OR UNSCHEDULED ASSESSMENTS OF SAFETY PARAMETERS

If a participant has repeated assessments before the start of the first treatment, the results from the final nonmissing assessment made prior to the start of the study intervention will be used as baseline. If end-of-study assessments are repeated or if unscheduled visits occur, the last nonmissing postbaseline assessment will be used as the end-of-study assessment for generating summary statistics.

15.6 MISSING SEVERITY ASSESSMENT FOR ADVERSE EVENTS

If severity is missing for an AE that started before the date of the first dose of study intervention, an intensity of *mild* will be assigned. If severity is missing for an AE that started on or after the date of the first dose of study intervention, an intensity of *severe* will be assigned. The imputed values for severity assessment will be used for the incidence summary; the values will be shown as missing in the data listings.

15.7 MISSING CAUSAL RELATIONSHIP TO STUDY INTERVENTION FOR ADVERSE EVENTS

If the causal relationship to the study intervention is missing for an AE that started on or after the date of the first dose of study intervention, a causality of *yes* will be assigned. The imputed values for causal relationship to study intervention will be used for the incidence summary; the values will be shown as missing in the data listings.

15.8 MISSING DATE INFORMATION FOR ADVERSE EVENTS

Partial adverse event onset date will be imputed as follows: 1) if day is missing but month is not, impute the date as the first day of the month; 2) if both day and month are missing, impute the date as 01 Jan; 3) if imputed onset date is before the first treatment, yet the corresponding adverse event was not observed pre-treatment, then impute the onset date as the first treatment date. Imputed partial adverse event onset date will only be used to determine the adverse event onset cycle.

Other partial adverse event dates will not be imputed. All partial dates will be listed "as is" in the data listings.

15.9 MISSING DATE INFORMATION FOR PRIOR OR CONCOMITANT MEDICATIONS

There will be no imputation for missing or partially reported medication start and/or end dates. However, reported partial information will be utilized in classification of prior and concomitant medications, when appropriate. For example, a medication may be classified as a pre-study medication if the partial end date provided is determined to definitively have occurred prior to the study intervention date (for example, the partial end date provided is "2017", and Day 1 occurred on August 29, 2018).

If start and/or stop dates for medications are only partially reported but can be classified as having occurred prior to Day 1, then the medications will be included in this summary of prior medications.

If stop dates for medications are only partially reported and cannot be definitively classified as having occurred prior to Day 1, then the medications will be included in this summary of concomitant medications.

[redacted]

17.0 REFERENCES

Dmitrienko A, Molenberghs G, Chuang-Stein C, Offen W. Analysis of Clinical Trials Using SAS: A Practical Guide. SAS Institute, Cary, North Carolina 2005; 104-108.

O'Kelly M, Ratitch B. Clinical trials with missing data: A guide for practitioner. Statistics in Practice. Wiley;2014.

Wilson EB, Hilferty, MM. The distribution of chi-squared. Proceedings of the National Academy of Sciences, Washington. 1931;17:12,684-688.

18.0 HISTORY OF CHANGES

Amendment 1:

Date	Section(s)	Description
[redacted]	[redacted]	[redacted]
11/18/2020	7.0	Deleted word “cumulative” to remove confusion.
[redacted]	[redacted]	[redacted]
11/18/2020	11.1	Added summary for AESI.
[redacted]	[redacted]	[redacted]
11/18/2020	15.4	Added detail about the calculation of [redacted] transformed total score.
[redacted]	[redacted]	[redacted]
11/18/2020	17.0	Added references.