STATISTICAL ANALYSIS PLAN ADDENDUM

Study Title:	A Randomized, Double-blind, Placebo- and Active-controlled, Multicenter, Phase 3 Study to Assess the Efficacy and Safety of Filgotinib Administered for 52 weeks in Combination with Methotrexate to Subjects with Moderately to Severely Active Rheumatoid Arthritis Who Have an Inadequate Response to Methotrexate
Name of Test Drug:	Filgotinib
Study Number:	GS-US-417-0301
Protocol Version (Date):	Amendment 1 (05 July 2016)
Analysis Type:	Final Analysis
Analysis Plan Version:	Version 2.0
Analysis Plan Date:	12 August 2019
Analysis Plan Author(s):	Lei Ye, Ph.D.

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

1. INTRODUCTION

This statistical analysis plan (SAP) addendum provides the additional efficacy analyses per the Food and Drug Administration (FDA) comments on the SAPs of Studies GS-US-417-0301 and GS-US-417-0303 dated 15 February 2019 (Reference ID: 4391728).

In addition, the following additional analysis through Week 52 will also be provided:

- Low disease activity (LDA) and Remission per Clinical Disease Activity Index (CDAI) and Simplified Disease Activity Index (SDAI) criteria
- Boolean remission
- Lymphocyte subset counts (TBNK cells)
- Immunoglobulins (Ig) A, G, M, and total Ig level

The detailed statistical methods and data presentations are defined in the original SAP of this study.

2. EFFICACY ANALYSES

2.1. General Considerations

Estimands:

Three efficacy estimands, composite estimand, treatment-policy estimand and hypothetical estimand are defined for the primary and key secondary efficacy endpoints, respectively, as described in Table 2-1, Table 2-2 and Table 2-3 below based on ICH E9R (1) {International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) 2017}.

Table 2-1. Estimands for Binary Endpoints: Filgotinib versus Placebo

Binary Endpoint	Composite	Treatment Policy	Hypothetical
Population	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)
Patient-level outcome to be measured	Outcome Response for a binary endpoint at Week 12 (e.g., ACR20 response at Week 12)	Outcome Response for a binary endpoint at Week 12 (e.g., ACR20 response at Week 12)	Outcome Response for a binary endpoint at Week 12 (e.g., ACR20 response at Week 12)
Measure of intervention effect & handling of intercurrent events.	Measure of the effect of the initially randomized treatments assuming that any rescue efforts (i.e., escape to standard of care) or premature discontinuations from treatment result in a non-response	Measure of the treatment effect regardless of what treatment was actually received and ignoring the occurrence of intercurrent events (i.e., protocol violations, escape to standard of care, or premature study discontinuation)	Measure of the effect of the initially randomized treatments assuming all patients had remained on their randomized treatment throughout the study (i.e., assuming patients did not receive rescue medication or did not discontinue study)
Population-level summary measure	Difference in a binary efficacy endpoint, comparing those assigned to filgotinib to those assigned to placebo	Difference in a binary efficacy endpoint, comparing those assigned to filgotinib to those assigned to placebo	Difference in a binary efficacy endpoint, comparing those assigned to filgotinib to those assigned to placebo
Estimators	**Main Estimator**: The treatment difference for a binary efficacy endpoint will be estimated using a logistic regression model. Missing data will be imputed using non-responder imputation.	**Main Estimator**: The treatment difference for a binary efficacy endpoint will be estimated using a logistic regression model. Missing data will be imputed using multiple imputation. **Sensitivity Estimator**: The tipping point method will be used. **Supportive Estimator**: Only observed data will be used for this model. If a subject discontinues from the study, the data will be treated as missing from the point of loss-to-follow-up to the end of the study. No imputation will be performed.	**Main Estimator**: The treatment difference for a binary efficacy endpoint will be estimated using a logistic regression model. Missing data will be imputed using multiple imputation (see bullet points below for additional details regarding missing data). If a subject takes standard of care medications due to inadequate response or discontinues from study treatment, the data will be treated as missing from the point of inadequate response or study drug discontinuation. If a subject discontinues from the study, treat data as missing from the point of loss-to-follow-up onward to the end of the study. **Sensitivity Estimator**: The tipping point method will be used.

Table 2-2. Estimands for Binary Endpoints: Filgotinib versus Adalimumab

Binary Endpoint	Composite	Treatment Policy	Hypothetical
Population	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)
Patient-level outcome to be measured	Outcome Response for a binary endpoint at Week 12 (e.g., DAS28(CRP) ≤ 3.2; DAS28(CRP) < 2.6)	Outcome Response for a binary endpoint at Week 12 (e.g., DAS28(CRP) ≤ 3.2; DAS28(CRP) < 2.6)	Outcome Response for a binary endpoint at Week 12 (e.g., DAS28(CRP) ≤ 3.2; DAS28(CRP) < 2.6)
Measure of intervention effect & handling of intercurrent events.	Measure of the effect of the initially randomized treatments assuming that any rescue efforts (i.e., escape to standard of care) or premature discontinuations from treatment result in a non-response	Measure of the treatment effect regardless of what treatment was actually received and ignoring the occurrence of intercurrent events (i.e., protocol violations, escape to standard of care, or premature study discontinuation)	Measure of the effect of the initially randomized treatments assuming all patients had remained on their randomized treatment throughout the study (i.e., assuming patients did not receive rescue medication or did not discontinue study)
Population-level summary measure	Difference in a binary efficacy endpoint, comparing those assigned to filgotinib to those assigned to adalimumab	Difference in a binary efficacy endpoint, comparing those assigned to filgotinib to those assigned to adalimumab	Difference in a binary efficacy endpoint, comparing those assigned to filgotinib to those assigned to adalimumab
Estimators	**Main Estimator**: The non-inferiority of filgotinib over adalimumab for a binary efficacy endpoint will be estimated using non-inferiority test. Missing data will be imputed using non-responder imputation.	**Main Estimator**: The non-inferiority of filgotinib over adalimumab for a binary efficacy endpoint will be estimated using non-inferiority test. Missing data will be imputed using multiple imputation. **Sensitivity Estimator**: The tipping point method will be used. **Supportive Estimator**: Only observed data will be used for this model. If a subject discontinues from the study, the data will be treated as missing from the point of loss-to-follow-up to the end of the study. No imputation will be performed.	**Main Estimator**: The non-inferiority of filgotinib over adalimumab for a binary efficacy endpoint will be estimated using non-inferiority test. Missing data will be imputed using multiple imputation (See bullet points below for additional details regarding missing data). If a subject takes standard of care medications due to inadequate response or discontinues from study treatment, the data will be treated as missing from the point of inadequate response or study drug discontinuation. If a subject discontinues from the study, treat data as missing from the point of loss-to-follow-up onward to the end of the study. **Sensitivity Estimator**: The tipping point method will be used.

Table 2-3. Estimands for Continuous Endpoints: Filgotinib versus Placebo

Continuous Endpoint	Composite	Treatment Policy	Hypothetical
Population	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)	Full Analysis Set as defined in the protocols (Sections 4.2, 4.3, and 8.2.1)
Patient-level outcome to be measured	Change from baseline in the continuous outcome at Week 12 or Week 24 (e.g., Change from baseline in mTSS at Week 24)	Change from baseline in the continuous outcome at Week 12 or Week 24 (e.g., Change from baseline in mTSS at Week 24)	Change from baseline in the continuous outcome at Week 12 or Week 24 (e.g., Change from baseline in mTSS at Week 24)
Measure of intervention effect & handling of intercurrent events	Measure of the effect of the initially randomized treatments assuming that any rescue efforts (i.e., escape to standard of care) or premature discontinuations from treatment result in an unfavorable outcome	Measure of the treatment effect regardless of what treatment was actually received and ignoring the occurrence of intercurrent events (i.e., protocol violations, escape to standard of care, or premature study discontinuation)	Measure of the effect of the initially randomized treatments assuming all patients had remained on their randomized treatment throughout the study (i.e., assuming patients did not receive rescue medication or did not discontinue study)
Population-level summary measure	Mean difference between treatment arms (comparing those assigned to filgotinib to those assigned to placebo) in the change from baseline to Week 12 or Week 24 in a given continuous outcome	Mean difference between treatment arms (comparing those assigned to filgotinib to those assigned to placebo) in the change from baseline to Week 12 or Week 24 in a given continuous outcome	Mean difference between treatment arms (comparing those assigned to filgotinib to those assigned to placebo) in the change from baseline to Week 12 or Week 24 in a given continuous outcome
Estimator(s)	**Main Estimator**: The treatment difference for a continuous efficacy endpoint will be estimated using a mixed effects model. Missing data will be imputed using multiple imputation based on observed data from the placebo group.	**Main Estimator**: The treatment difference for a continuous efficacy endpoint will be estimated using a mixed effects model. Missing data will be imputed using multiple imputation . **Sensitivity Estimator**: The tipping point method will be used. **Supportive Estimator**: An ANCOVA model using observed data only will be fit. If a subject discontinues from the study, the data will be treated as missing from the point of loss-to-follow-up to the end of the study. No imputation will be performed.	**Main Estimator**: The treatment difference for a continuous efficacy endpoint will be estimated using a mixed effects model. Missing data will be imputed using multiple imputation (see bullet points below for additional details regarding missing data). If a subject takes standard of care medications due to inadequate response or discontinues from study treatment, the data will be treated as missing from the point of inadequate response or study drug discontinuation. If a subject discontinues from the study, the data will be treated as missing from the point of loss-to-follow-up to the end of the study. **Sensitivity Estimator**: The tipping point method will be used.

2.2. Additional Analyses of Efficacy Endpoints per FDA Comments

The following additional analyses will be provided for estimands:

- For composite estimand for continuous endpoints, the main estimator will be performed using the multiple imputation approach with missing data being imputed based on data observed from the placebo group.

  — For key continuous endpoints (change from baseline in HAQ-DI, SF-36 PCS, FACIT-F at Week 12 and mTSS at Week 24), multiple imputation will be performed by generating the multiple imputed datasets based on linear regression models on observed scores from the placebo group. These multiple imputed datasets are then analyzed by using the analysis method specified in the original SAP Section 6.3.2 for complete data. The results from each set of imputed datasets will then be combined using Rubin's rule {Rubin 1987}.

- For treatment policy estimand, the supportive estimator will be performed based on an ANCOVA model using only observed data:

  — For key continuous efficacy endpoints (change from baseline in HAQ-DI, SF-36 PCS, FACIT-F at Week 12 and mTSS at Week 24), an ANCOVA model with baseline value, stratification factors and treatment group in the model will be used for comparing a treatment difference between each filgotinib group and placebo in change from baseline to the time point of interest. The LS means along with the two-sided 95% CIs and the p-value for the difference in mean change from Baseline at the time point of interest from the model will be presented. The observed values will be used for analysis. If a subject discontinues from the study, the data will be treated as missing from the point of loss-to-follow-up to the end of the study. No imputation will be performed.

The following supportive analyses will be provided based on the FDA comments:

- Per the FDA comment that the MMRM model should also include subjects with baseline measurement available, a revised MMRM model will be performed for key continuous efficacy endpoints (change from baseline in HAQ-DI, SF-36 PCS, FACIT-F at Week 12 and mTSS at Week 24). The model will include all the subject with baseline measurement with stratification factors, treatment, visit, and treatment by visit interaction included as fixed effects and subject as the random effect. The model will use actual observations including baseline as outcome and the baseline will not be adjusted as covariate. This analysis will be based on the on-treatment data and all available data as described in Section 6.1 of the original SAP.

- Per the FDA comment on the analysis window, the following changes will be made on the window definition: for mTSS and PRO endpoints (SF-36 PCS, and FACIT-F), visit windows will include +/- 4 weeks as described in Table 2-4 and Table 2-5. The analysis will follow the same analysis method specified for primary analysis of these endpoints as described in Section 6.3.2 of the original SAP.

Table 2-4. Analysis Visit Windows for On-treatment mTSS Data

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Baseline	1	(none)	15
Week 12	85	58	113
Week 24	169	142	197

Table 2-5. Analysis Visit Windows for On-treatment SF-36 and FACIT-Fatigue

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 4	29	2	57
Week 12	85	58	113
Week 24	169	142	197

2.3. Additional Analyses of Low Disease Activity and Remission

The following additional efficacy endpoints will be analyzed:

- Low disease activity (LDA) per CDAI or SDAI criteria is defined as CDAI ≤ 10 or SDAI ≤ 11.
- Remission per CDAI or SDAI criteria is defined as CDAI ≤ 2.8 or SDAI ≤ 3.3.
- Boolean remission is achieved when tender joint count 28 (TJC28) ≤ 1 and swollen joint count 28 (SJC28) ≤ 1 and c-reactive protein (CRP) ≤ 1 mg/dL and subject's global assessment (SGA) ≤ 1 (on a 0-10 cm scale). Boolean remission is not achieved when at least 1 of the 4 components has a value > 1. In case that some components are missing, Boolean remission is not achieved if at least 1 of the non-missing components has a value > 1. If non-missing components are not sufficient to determine Boolean remission, then Boolean remission will be considered as missing.

The proportion of subjects who achieved LDA or Remission per the above criteria will be analyzed using main estimator of treatment policy estimand and composite estimand, respectively. In addition, the proportion of subjects who achieved LDA or Remission will be analyzed through Week 24 using the logistic regression method with NRI.

3. ADDITIONAL SAFETY ANALYSES

The descriptive statistics of Baseline values, values at each postbaseline visit and change from Baseline at each postbaseline visit will be provided by treatment groups for the following laboratory tests.

- Lymphocyte subset counts (TBNK cells)
  - CD3 (%)
  - CD3 (/uL)
  - CD4 (%)
  - CD4 (/uL)
  - CD8 (%)
  - CD8 (/uL)
  - CD19 (%)
  - CD19 (/uL)
  - CD16+56 (%)
  - CD16+56 (/uL)
- Immunoglobulins (Ig) A, G, M, and total Ig level

In addition, the proportion of subjects below the lower limit of normal (LLN) or above the upper limit of normal (ULN) will be summarized descriptively for TBNK cells. The proportion of subjects below the LLN will be summarized descriptively for Immunoglobulins (Ig) A, G, and M.

The list of tables, figures, and listings (TFLs) for the additional analyses is shown in Appendix 1.

4. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
12 August 2019	Section 2.1	Removal of treatment policy main estimator as the primary analysis and added as supportive analysis	Based on the Division's feedback at the 07 August 2019 Type C meeting, the pre-specified analysis will remain as the primary analysis.

5. REFERENCE

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. Current Step 1 Version. Dated 16 June. 2017.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons, Inc; 1987.

6. APPENDICES

Appendix 1. Tables, Figures, and Listings (TFLs) for the Additional Analyses

Table Number	Title
1	HAQ-DI: Change from Baseline at Week 12 - Multiple Imputation with Unfavorable Outcome
2	HAQ-DI: Change from Baseline at Week 12 - All Available Data Included (ANCOVA analysis)
3	HAQ-DI: Change from Baseline at Week 12 - MMRM including Subjects with Baseline Value Available
4	HAQ-DI: Change from Baseline at Week 12 - All Available Data Included (MMRM including Subjects with Baseline Value Available)
5	van der Heijde modified Total Sharp Score (mTSS): Change from Baseline at Week 24 - Multiple Imputation with Unfavorable Outcome
6	van der Heijde modified Total Sharp Score (mTSS): Change from Baseline at Week 24 - All Available Data Included (ANCOVA analysis)
7	van der Heijde modified Total Sharp Score (mTSS): Change from Baseline at Week 24 - MMRM including Subjects with Baseline Value Available
8	van der Heijde modified Total Sharp Score (mTSS): Change from Baseline at Week 24 - All Available Data Included (MMRM including Subjects with Baseline Value Available)
9	SF-36 Physical Component Summary (PCS): Change from Baseline at Week 12 - Multiple Imputation with Unfavorable Outcome
10	SF-36 Physical Component Summary (PCS): Change from Baseline at Week 12 - All Available Data Included (ANCOVA analysis)
11	SF-36 Physical Component Summary (PCS): Change from Baseline at Week 12 - MMRM including Subjects with Baseline Value Available
12	SF-36 Physical Component Summary (PCS): Change from Baseline at Week 12 - All Available Data Included (MMRM including Subjects with Baseline Value Available)
13	FACIT-Fatigue: Change from Baseline at Week 12 - Multiple Imputation with Unfavorable Outcome
14	FACIT-Fatigue: Change from Baseline at Week 12 - All Available Data Included (ANCOVA analysis)
15	FACIT-Fatigue: Change from Baseline at Week 12 - MMRM including Subjects with Baseline Value Available
16	FACIT-Fatigue: Change from Baseline at Week 12 - All Available Data Included (MMRM including Subjects with Baseline Value Available)
17	van der Heijde modified Total Sharp Score (mTSS): Change from Baseline at Week 24 – Revised Analysis Window
18	van der Heijde modified Total Sharp Score (mTSS): Change from Baseline at Week 24 - All Available Data Included (Revised Analysis Window)
19	SF-36 Physical Component Summary (PCS): Change from Baseline at Week 12 - Revised Analysis Window
20	SF-36 Physical Component Summary (PCS): Change from Baseline at Week 12 - All Available Data Included (Revised Analysis Window)

Table Number	Title
21	FACIT-Fatigue: Change from Baseline at Week 12 - Revised Analysis Window
22	FACIT-Fatigue: Change from Baseline at Week 12 - All Available Data Included (Revised Analysis Window)
23	Clinical Disease Activity Index (CDAI) <= 10 by Visit through Week 24: Pairwise Comparisons versus Placebo (Logistic Regression with NRI)
24	Clinical Disease Activity Index (CDAI) <= 10 by Visit from Week 24 to Week 52 - All Available Data Included
25	Clinical Disease Activity Index (CDAI) <= 2.8 by Visit through Week 24: Pairwise Comparisons versus Placebo (Logistic Regression with NRI)
26	Clinical Disease Activity Index (CDAI) <= 2.8 by Visit from Week 24 to Week 52- All Available Data Included
27	Simplified Disease Activity Index (SDAI) <= 11 by Visit through Week 24 : Pairwise Comparisons versus Placebo (Logistic Regression with NRI)
28	Simplified Disease Activity Index (SDAI) <= 11 by Visit from Week 24 to Week 52 - All Available Data Included
29	Simplified Disease Activity Index (SDAI) <= 3.3 by Visit through Week 24: Pairwise Comparisons versus Placebo (Logistic Regression with NRI)
30	Simplified Disease Activity Index (SDAI) <= 3.3 by Visit from Week 24 to Week 52- All Available Data Included
31	Boolean Remission by Visit through Week 24: Pairwise Comparisons versus Placebo (Logistic Regression with NRI)
32	Boolean Remission by Visit from Week 24 to Week 52 - All Available Data Included
33	Clinical Disease Activity Index (CDAI) <= 2.8 at Week 12: Pairwise Comparisons versus Placebo - All Available Data Included (Logistic Regression with Multiple Imputation)
34	Clinical Disease Activity Index (CDAI) <= 10 at Week 12: Pairwise Comparisons versus Placebo - All Available Data Included (Logistic Regression with Multiple Imputation)
35	Simplified Disease Activity Index (SDAI) <= 3.3 at Week 12: Pairwise Comparisons versus Placebo - All Available Data Included (Logistic Regression with Multiple Imputation)
36	Simplified Disease Activity Index (SDAI) <= 11 at Week 12: Pairwise Comparisons versus Placebo - All Available Data Included (Logistic Regression with Multiple Imputation)
37	Boolean Remission at Week 12: Pairwise Comparisons versus Placebo - All Available Data Included (Logistic Regression with Multiple Imputation)
38	Cellmark: TBNK cell parameter Laboratory Events by Visit
39	Cellmark: TBNK cell parameter and Change from Baseline
40	Chemistry: Immunoglobulin A (mg/dL) Laboratory Events by Visit
41	Chemistry: Immunoglobulin G (mg/dL) Laboratory Events by Visit
42	Chemistry: Immunoglobulin M (mg/dL) Laboratory Events by Visit
43	Chemistry: Immunoglobulin A (mg/dL) and Change from Baseline
44	Chemistry: Immunoglobulin G (mg/dL) and Change from Baseline
45	Chemistry: Immunoglobulin M (mg/dL) and Change from Baseline
46	Chemistry: Total Immunoglobulin (mg/dL) and Change from Baseline