# 109MS303 / NCT00835770

## Biogen - BG00012 in MS

## Statistical Analysis Plan for Clinical Study Report for Submission


**BIOGEN**

**STATISTICAL ANALYSIS PLAN**

A Dose-Blind, Multicenter, Extension Study to Determine the Long-Term Safety and Efficacy of Two Doses of BG00012 Monotherapy in Subjects with Relapsing-Remitting Multiple Sclerosis

# Protocol 109MS303 Statistical Analysis Plan for Clinical Study Report for Submission

Study Phase: 3b
Product Studied: BG00012
Date of Protocol: 05 February 2018 (Version 8)
Date of SAP: 14 JUN 2019 (Final)

**Authored By:** [redacted]

[redacted] PhD [redacted] — 14 JUN 2019 Date

**Approved By:** [redacted]

[redacted], PharmD, MPH [redacted] — 14 Jun 2019 Date

[redacted] — 14 JUN 2019 Date

[redacted], Ph.D. [redacted]

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP)

**Confidentiality Statement**

# APPROVAL

| This document has been reviewed and approved by: | | |
|---|---|---|
| [redacted] | [redacted] | 14 JUN 2019 |
| SMT Statistician | Signature | Date |
| [redacted] | [redacted] | 14 JUN 2019 |
| CDT Statistician (Printed Name) | Signature | Date |
| [redacted] | [redacted] | 14 Jun 2019 |
| SMT [redacted] (Printed Name) | Signature | Date |

TABLE OF CONTENTS

## List of Abbreviations

| | |
|---|---|
| AE | adverse event |
| ALT (SGPT) | alanine transaminase (serum glutamate pyruvate transaminase) |
| AST (SGOT) | aspartate transaminase (serum glutamic oxaloacetic transaminase) |
| ARR | annualized relapse rate |
| BUN | blood urea nitrogen |
| CRF | case report form |
| Gd | gadolinium |
| GGT | gamma-glutamyl-transferase |
| EDSS | Expanded Disability Status Scale |
| EQ-5D | EQ-5D Health Survey |
| GA | glatiramer acetate |
| HDL | high-density lipoprotein |
| HLGT | high-level group term |
| HLT | high-level term |
| IXRS | Interactive Voice and Web Response System |
| INEC | Independent Neurology Evaluation Committee |
| LDL | low-density lipoprotein |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MRI | magnetic resonance imaging |
| MS | multiple sclerosis |
| MTR | Magnetization Transfer Ratio |
| PBVC | Percent Brain Volume Change |
| PT | preferred term |
| SAE | serious adverse event |
| SAP | Statistical Analysis Plan |
| SOC | system organ class |
| SF-36 | SF-36® Health Survey |
| ULN | upper limit of normal |
| VAS | Visual Analogue Scale |

| | |
|---|---|
| VFT | Visual Function Test |
| WBC | white blood cell |
| WHO | World Health Organization |

# 1 STUDY OBJECTIVES AND ENDPOINTS

## 1.1 Primary Objective and Endpoint

The primary objective of the study is to evaluate the long-term safety profile of BG00012.

## 1.2 Secondary Objectives and Endpoints

The secondary objectives of this study are as follows:

- To evaluate the long-term efficacy of BG00012 using clinical endpoints (including relapse and ARR) and disability progression (Expanded Disability Status Scale [EDSS]).
- To evaluate further the long-term effects of BG00012 on MS brain lesions on MRI scans in subjects who had MRI scans as part of Studies 109MS301 and 109MS302.
- To evaluate the long-term effects of BG00012 on health economics assessments and the visual function test (VFT). The endpoints are the Short-form 36 Health Survey (SF-36) and EuroQol EQ-5D Health Survey (EQ-5D) quality of life questionnaires, and the visual function test scores.

## 1.3 Exploratory Objectives and Endpoints

The exploratory objectives of this study are as follows:

- [redacted]
- [redacted]

# 2 STUDY DESIGN

## 2.1 Study Overview

The first phase of this extension study is a multicenter, parallel-group, randomized, dose-blind, dose-comparison study. The second phase of the study is an open-label phase during which all subjects receive BG00012 at a dose of 240mg BID. A total of 1738 subjects with relapsing-remitting multiple sclerosis have been randomized at approximately 375 sites in North America, Europe, and the rest of the world.

Eligible subjects from Studies 109MS301 and 109MS302 were enrolled in this extension study and were followed for up to 8 years, assuming BG00012 continues to have a positive benefit/risk ratio.

All subjects received either:

• BG00012 240 mg BID (2 capsules [120 mg each] twice a day and 2 placebo capsules once a day); or

• BG00012 240 mg TID (2 capsules [120 mg each] 3 times a day).

Subjects randomized to BG00012 in Study 109MS301 or 109MS302 continued on the same BG000012 dose in this extension study.

Subjects randomized to placebo in Study 109MS301 or 109MS302 or glatiramer acetate (GA) in Study 109MS302 were randomized to BG00012 in a 1:1 ratio to 240 mg BID or 240 mg TID in this extension study.

With regulatory approval (effective March 2013), all patients receiving BG00012 240 mg TID switched to BID (approved dosage) at their next study visit. This study became an open label extension study with the approved dosage BG00012 240 mg BID.

Subjects who switched to an alternate approved MS therapy in Study 109MS301 or 109MS302 were randomized/enrolled in this extension study as outlined above according to their original treatment group in 109MS301 or 109MS302.

## 2.2 Overall Study Duration and Follow-Up

Eligible subjects were enrolled at Week 96 (Visit 24) of their previous BG00012 study (109MS301 or 109MS302), which would serve as the Baseline Visit for this extension study. If the Baseline Visit could not be combined with Visit 24 of their previous BG00012 study, the subjects may have been randomized within 6 months of Visit 24, provided that they still were eligible for enrollment. The Baseline visit assessments may have been repeated if there was a gap in study treatment.

Subjects reported to the study site every 4 weeks for the first 24 weeks (Visit 1 through Visit 6) and every 12 weeks thereafter for up to 8 years.

Subjects who completed the Week 384 visit (Visit 36), regardless of further need to monitor lymphocytes on study, completed an End of Study Visit at Week 388 ± 7 days. For subjects who have completed Visit 36, an End of Study Visit was completed 4 weeks after their next visit (scheduled or unscheduled), regardless of further need to monitor lymphocytes on study. Subjects who permanently discontinued BG00012 prior to Week 384, regardless of further need to monitor lymphocytes on study, completed a Premature Study Treatment Withdrawal Visit within 4 weeks after their last dose of BG00012. Subjects who completed BG00012 treatment or permanently discontinued BG00012 for any reason and who had a lymphocyte count <LLN were not withdrawn from the study. These subjects were followed according to Protocol Section 11.3.

For subjects who have already completed Visit 36 (or beyond) prior to the Protocol Version 8, an End of Study Visit was completed 4 weeks after their next visit (scheduled or unscheduled), regardless of further need to monitor lymphocytes on study. For subjects who have not completed Visit 36 prior to Protocol Version 8, Visit 36 served as the last scheduled Visit and an End of Study Visit was completed 4 weeks after their next visit (scheduled or unscheduled), regardless of further need to monitor lymphocytes on study.

Premature Study Withdrawal Visits and Unscheduled Relapse Assessment Visits were performed as necessary. All efforts were made to keep the subjects' visit schedule relative to their Baseline Visit (Day 1). All subjects who discontinued the study treatment were considered to be prematurely withdrawn from the study.

The end of study is defined as the Last Patient, Last Visit for final collection of data for the primary outcome

### 2.3 MRI

MRI was performed in the same subset of sites/countries as in Studies 109MS301 and 109MS302, which was based on the availability of MRI equipment at the sites to perform the assessments required for the study (e.g., MTR). All subjects at the MRI sites were offered the option to participate in Studies 109MS301 and 109MS302, and those subjects continuing into Study 109MS303 continued their participation in the MRI cohort. MRI scans were performed on a yearly basis in Study 109MS303, as well as at the time of premature withdrawal up through and including Amendment 6.

## 3 STUDY ACTIVITIES

The study flowchart for the current version of the protocol is presented below.

## 3.1 Study Activities (Year 1 Visits)

| Tests and Assessments[1] | Baseline Visit (Day 1)[2] | Visit 1 (Week 4 ±5 days) | Visit 2 (Week 8 ±5 days) | Visit 3 (Week 12 ±5 days) | Visit 4 (Week 16 ±5 days) | Visit 5 (Week 20 ±5 days) | Visit 6 (Week 24 ±5 days) | Visit 7 (Week 36 ±7 days) | Visit 8 (Week 48 ±7 days) |
|---|---|---|---|---|---|---|---|---|---|
| Informed Consent | X | | | | | | | | |
| Randomization | X | | | | | | | | |
| Medical History Update | X | | | | | | | | |
| Physical Examination | X | | | | | | X | | X |
| Vital Signs | X | X | X | X | X | X | X | X | X |
| Hematology | X | | | X | | | X | X | X |
| Blood Chemistry | X | X | X | X | X | X | X | X | X |
| Urine Pregnancy Test[3] | X | X | X | X | X | X | X | X | X |
| Urinalysis | X | X[4] | X[4] | X[4] | X[4] | X[4] | X[4] | X[4] | X[4] |
| Brain MRI[5] Scan ± Gd[6] | Scheduled MRI scans were performed in protocol versions 1 through 6. | | | | | | | | |
| EDSS | X | | | X[7] | | | X | | X |
| Visual Function Test | X | | | | | | X | | X |
| SF-36 and EQ-5D | X | | | | | | | | X |
| Dispense Study Treatment (BG00012) | X | X | X | X | X | X | X | X | X |
| Concomitant Therapy and Procedures Recording | Monitor and record throughout the study as per Protocol Section 11.6 | | | | | | | | |
| AE/SAE Reporting | Monitor and record throughout the study as per Protocol Section 15 | | | | | | | | |

AE = adverse event; EDSS = Expanded Disability Status Scale; EQ-5D = EuroQol EQ-5D Health Survey; Gd = gadolinium; MRI = magnetic resonance imaging; SAE = serious adverse event; SF-36 = Short-form 36 Health Survey.

1. Tests and assessments must be completed prior to BG00012 distribution.
2. The Week 96 visit (Visit 24) for 109MS301 and 109MS302 will serve as the Baseline Visit for this extension study; activities should be completed (where appropriate) by staff of the parent study.
3. For women of childbearing potential. Results must be known prior to BG00012 distribution.
4. If urinalysis is abnormal, see Protocol Section 11.2.
5. Only in subjects who had MRI scans as part of Studies 109MS301 and 109MS302 and in 109MS303 up through and including Amendment 6.
6. MRI must not be performed within 30 days of receiving a course of steroids.
7. rmed only for subjects who at the Week 96 visit (Visit 24) in 109MS301 or 109MS302 had a 1.0 point increase in their EDSS score from a baseline EDSS 1.0 or a 1.5 point increase in their EDSS from a baseline EDSS = 0.

## 3.2 Study Activities (Years 2 and 3 Treatment Visits)

| | Year 2 | | | | Year 3 | | | |
|---|---|---|---|---|---|---|---|---|
| **Visit** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** |
| **Study Week (±7 days)** | **60** | **72** | **84** | **96** | **108** | **120** | **132** | **144** |
| **Tests and Assessments**[1] | | | | | | | | |
| Physical Examination | X | X | X | X | X | X | X | X |
| Vital Signs | X | X | X | X | X | X | X | X |
| Hematology | X | X | X | X | X | X | X | X |
| Blood Chemistry | X | X | X | X | X | X | X | X |
| Urine Pregnancy Test[2] | X | X | X | X | X | X | X | X |
| Urinalysis | X[3] | X[3] | X[3] | X[3] | X[3] | X[3] | X[3] | X[3] |
| Brain MRI[4] Gd[5] | Scheduled MRI scans were performed in protocol versions 1 through 6. | | | | | | | |
| EDSS | X | X | X | X | X | X | X | X |
| Visual Function Test | X | X | X | X | X | X | X | X |
| SF-36 and EQ-5D | | | | X | | | | X |
| Dispense Study Treatment (BG0012) | X | X | X | X | X | X | X | X |
| Concomitant Therapy and Procedures Recording | Monitor and record throughout the study as per Protocol Section 11.6 | | | | | | | |
| AE/SAE Reporting | Monitor and record throughout the study as per Protocol Section 15 | | | | | | | |

AE = adverse event; EDSS = Expanded Disability Status Scale; EQ-5D = EuroQol EQ-5D Health Survey; Gd = gadolinium; MRI = magnetic resonance imaging; SAE = serious adverse event; SF-36 = Short-form 36 Health Survey.

1. Tests and assessments must be completed prior to BG00012 distribution.
2. For women of childbearing potential. Results must be known prior to BG00012 distribution.
3. If urinalysis is abnormal, see Protocol Section 11.2.
4. Only in subjects who had MRI scans as part of Studies 109MS301 and 109MS302 and in 109MS303 up through and including Amendment 6.
5. MRI must not be performed within 30 days of receiving a course of steroids.

## 3.3 Study Activities (Years 4 Through 8 Treatment Visits)

| Year | Year 4 | | | | Year 5 | | | | Year 6 | | | | Year 7 | | | | Year 8 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Visit** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36[1]** |
| **Study Week (±7 days)** | **156** | **168** | **180** | **192** | **204** | **216** | **228** | **240** | **252** | **264** | **276** | **288** | **300** | **312** | **324** | **336** | **348** | **360** | **372** | **384** |
| **Tests and Assessments[2]** | | | | | | | | | | | | | | | | | | | | |
| Physical Examination | | X | | X | | X | | X | | X | | X | | X | | X | | X | | X |
| Vital Signs | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Hematology | See Section 3.5 | | | | | | | | | | | | | | | | | | | |
| Blood Chemistry | | X | | X | | X | | X | | X | | X | | X | | X | | X | | X |
| [redacted] | | | | | | | | | | | | | | | | | | | | |
| Urine Pregnancy Test[3] | | X | | X | | X | | X | | X | | X | | X | | X | | X | | X |
| Urinalysis | | X[4] | | X[4] | | X[4] | | X[4] | | X[4] | | X[4] | | X[4] | | X[4] | | X[4] | | X[4] |
| Brain MRI[5] Scan ± Gd[6] | Scheduled MRI scans were performed in protocol versions 1 through 6. | | | | | | | | | | | | | | | | | | | |
| EDSS | | X | | X | | X | | X | | X | | X | | X | | X | | X | | X |
| Visual Function Test | | X | | X | | X | | X | | X | | X | | X | | X | | X | | X |
| SF-36 and EQ-5D | | | | X | | | | X | | | | X | | | | X | | | | X |
| Dispense Study Treatment (BG00012)[7] | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | |
| Concomitant Therapy and Procedures Recording | Monitor and record throughout the study as per Protocol Section 11.6 | | | | | | | | | | | | | | | | | | | |
| AE/SAE Reporting | Monitor and record throughout the study as per Protocol Section 15 | | | | | | | | | | | | | | | | | | | |

AE = adverse event; CRF = case report form; [redacted]; EDSS = Expanded Disability Status Scale; EQ-5D = EuroQol EQ-5D Health Survey; Gd = gadolinium; MRI = magnetic resonance imaging; [redacted]; SAE = serious adverse event; SF-36 = Short-form 36 Health Survey.

1. For subjects who have already completed Visit 36 (or beyond), an End of Study Visit should be completed 4 weeks after their next visit (scheduled or unscheduled), regardless of further need to monitor lymphocytes on study.
2. Tests and assessments must be completed prior to BG00012 distribution.
3. For women of childbearing potential. Results must be known prior to BG00012 distribution.
4. If urinalysis is abnormal, see Protocol Section 11.2.

5. Only in subjects who had MRI scans as part of Studies 109MS301 and 109MS302 and in 109MS303 up through and including Amendment 6.
6. MRI must not be performed within 30 days of receiving a course of steroids.
7. The date and Study Visit/Week at which subjects are switched to open-label treatment should be recorded on the dosing CRF.

## 3.4 Study Activities for Subjects Who Already Completed Visit 36

**For subjects who have already completed Visit 36 (or beyond) prior to Protocol Version 8, an End of Study Visit should be completed 4 weeks after their next visit (scheduled or unscheduled), regardless of further need to monitor lymphocytes on study. For subjects who have not completed Visit 36 prior to Protocol Version 8, Visit 36 will serve as the last scheduled Visit and an End of Study Visit should be completed 4 weeks after their next visit (scheduled or unscheduled), regardless of further need to monitor lymphocytes on study.**

| Year | Year 9 | | | | Year 10 | | | | Year 11 | | | | Year 12 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Visit | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 |
| Study Week ±7 days) | 396 | 408 | 420 | 432 | 444 | 456 | 468 | 480 | 492 | 504 | 516 | 528 | 540 | 552 | 564 | 576 |
| **Tests and Assessments**[1] | | | | | | | | | | | | | | | | |
| Physical Examination | | X | | X | | X | | X | | X | | X | | X | | X |
| Vital Signs | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Hematology | See Protocol Section 4.5 | | | | | | | | | | | | | | | |
| Blood Chemistry | | X | | X | | X | | X | | X | | X | | X | | X |
| [redacted] | | | | | | | | | | | | | | | | |
| Urine Pregnancy Test[2] | | X | | X | | X | | X | | X | | X | | X | | X |
| Urinalysis | | X[3] | | X[3] | | X[3] | | X[3] | | X[3] | | X[3] | | X[3] | | X[3] |
| Brain MRI[4] Scan ± Gd[5] | Scheduled MRI scans were performed in protocol versions 1 through 6. | | | | | | | | | | | | | | | |
| EDSS | | X | | X | | X | | X | | X | | X | | X | | X |
| Visual Function Test | | X | | X | | X | | X | | X | | X | | X | | X |
| SF-36 and EQ-5D | | | | X | | | | X | | | | X | | | | X |
| Dispense Study Treatment (BG00012)[76] | | | | | | | | | | | | | | | | |
| Concomitant Therapy and Procedures Recording | Monitor and record throughout the study as per Protocol Section 11.6 | | | | | | | | | | | | | | | |
| AE/SAE Reporting | Monitor and record throughout the study as per Protocol Section 15 | | | | | | | | | | | | | | | |

AE = adverse event; CRF = case report form; [redacted]; EDSS = Expanded Disability Status Scale; EQ-5D = EuroQol EQ-5D Health Survey; Gd =

gadolinium; MRI = magnetic resonance imaging; [redacted]; SAE = serious adverse event; SF-36 = Short-form 36 Health Survey.

1. Tests and assessments must be completed prior to BG00012 distribution.
2. For women of childbearing potential. Results must be known prior to BG00012 distribution.
3. If urinalysis is abnormal, see Protocol Section 11.2.
4. Only in subjects who had MRI scans as part of Studies 109MS301 and 109MS302 and in 109MS303 up through and including Amendment 6.
5. MRI must not be performed within 30 days of receiving a course of steroids.
6. The date and Study Visit/Week at which subjects are switched to open-label treatment should be recorded on the dosing CRF.

### 3.5 Study Activities (Hematology and [redacted] Analysis: Year 4 Through End of Treatment Visits)

| Tests and Assessments | Subjects With Lymphocyte Count ≥LLN | Subjects With Lymphocyte Count <LLN to ≥500/mm$^3$ | Subjects With Lymphocyte Count <500/mm$^3$ |
|---|---|---|---|
| Hematology | Every 12 weeks | Every 12 weeks | Every 4 or 12 weeks per Protocol Table 3 |
| [redacted] | | | |
| Bone Marrow Biopsy (optional) | Not applicable | Not applicable | Optional procedure[6] |

[redacted]; LLN = lower limit of normal; [redacted].

[redacted]

6. See Protocol Section 11.3 for further details

## 3.6 Study Activities 109MS303 (Unscheduled and Post-Treatment Visits)

| Tests and Assessments | End of Study Visit[1] | Unscheduled Relapse Assessment Visit[2] | Premature Study Treatment Withdrawal Visit[3] | Lymphocyte Follow-Up (Up to 96 Weeks After Last Dose)[4] |
|---|---|---|---|---|
| Physical Examination | | X | X | |
| Vital Signs | X | X | X | X |
| Hematology | X[3] | X | X[3] | X |
| Blood Chemistry | X | X | X | |
| Urine Pregnancy Test[5] | X | X | X | |
| Urinalysis | X[6] | X[6] | X[6] | |
| Brain MRI[7] Scan ± Gd[8] | Scheduled MRI scans were performed in protocol versions 1 through 6. | | | |
| EDSS | | X | X | |
| Visual Function Test | | X | X | |
| SF-36 and EQ-5D | | X | X | |
| Relapse Assessment | | X | | |
| Concomitant Therapy and Procedures Recording | Monitor and record throughout the study as per Protocol Section 11.6 | | | |
| AE/SAE Reporting | Monitor and record throughout the study as per Protocol Section 15 | | | |

AE = adverse event; EDSS = Expanded Disability Status Scale; EQ-5D = EuroQol EQ-5D Health Survey; Gd = gadolinium; LLN = lower limit of normal; MRI = magnetic resonance imaging; SAE = serious adverse event; SF-36 = Short-form 36 Health Survey.

1. For subjects who have already completed Visit 36 or beyond, an End of Study Visit should be completed 4 weeks after their next visit (scheduled or unscheduled).
2. Unscheduled Relapse Assessment Visit is to be carried out within 5 days of suspected relapse.
3. Subjects who permanently discontinue BG00012 prior to Week 384 (Visit 36) should complete a Premature Study Treatment Withdrawal Visit within 4 weeks of their last dose. The Premature Study Treatment Withdrawal Visit will serve as the End of Study Visit.
4. Lymphocyte follow-up as per Protocol Section 11.3.
5. For women of childbearing potential.
6. If urinalysis is abnormal, see Protocol Section 11.2.
7. Only in subjects who had MRI scans as part of Studies 109MS301 and 109MS302 and in 109MS303 up through and including Amendment 6.
8. MRI must not be performed within 30 days of receiving a course of steroid.

# 4 INTERIM ANALYSIS

This statistical analysis plan (SAP) is intended to describe the planned analysis methods and considerations for the final analysis based on data after the database lock. Interim analyses were performed as needed for regulatory reporting, safety updates, publications, or as otherwise required by the Sponsor. One specific interim analysis was conducted to provide an assessment of the long-term safety and efficacy profile of BG00012 at the time of filing the MS application. This interim analysis was based on all data collected for subject visits, assessments, and evaluations through August 3, 2011. After that, several interim analyses were conducted based on data snapshots without formal database lock, and these results were presented for publications purposes to assess long-term safety and efficacy profile of BG00012.

# 5 SAMPLE SIZE JUSTIFICATION

There is no formal sample size calculation. The number of subjects eligible for this study is determined by the number of subjects who participated in and completed Studies 109MS301 and 109MS302.

Subjects randomized to BG00012 in Study 109MS301 or 109MS302 continued on the same BG000012 dose in this extension study. Subjects randomized to placebo in Study 109MS301 or 109MS302 or glatiramer acetate in Study 109MS302 were randomized to BG00012 in a 1:1 ratio to 240 mg BID or 240 mg TID in this extension study. Randomization was stratified by previous study and previous treatment (i.e., 3 strata: Study 109MS301 placebo, Study 109MS302 placebo, and Study 109MS302 glatiramer acetate).

All patients receiving BG00012 240 mg TID were switched to BID (approved dosage) at their next study visit after regulatory approval (effective March 2013).

# 6 STATISTICAL ANALYSIS METHODS

Descriptive summary statistics will be used throughout. For continuous endpoints and parameters, summary statistics will generally include: number of subjects with data, mean, standard deviation, median, and range. For categorical endpoints and parameters, summary statistics will generally include: number of subjects with data and the number and percentage of those with data in each category.

Summaries will generally include all available data included in the database lock for this final analysis. However, as this is an ongoing study, caution must be exercised when interpreting analyses at later timepoints for which sample sizes may be relatively sparse. Kaplan-Meier curves will display results through timepoints for which a sufficient number of subjects at risk exists in each treatment group (e.g., at least 20 subjects at risk per treatment group).

The statistical software, SAS® ,will be used for all statistical analyses. The SAS code for efficacy analyses is not provided in this statistical analysis plan, as the code is similar to that provided in the statistical analysis plans for Studies 109MS301 and 109MS302.

## 6.1 Statistical testing procedures

As this final analysis of a long-term extension study, analyses will generally be descriptive in nature and will be based on observed data, without imputation for missing values. For relevant efficacy analyses, confidence intervals will be provided for each treatment group where appropriate to characterize the variability around the point estimate. Statistical hypothesis testing may be completed for selected analyses and will be considered exploratory.

*Missing measurements*

In general, no imputation for missing measurements will be performed, unless specified otherwise.

*Missing dates*

For start date of adverse events, if all available start date information is the same as that of the date of first dose in 109MS303, and the event did not end prior to first dose, the date of first dose in 109MS303 will be used as the imputed date. Otherwise for adverse events, and always for the start of concomitant medications, the earliest possible start date based on available partial information will be used. Namely, if only the day is missing, the 1st of the month will be used; if the month is also missing, January 1st will be used. End dates for concomitant medications and adverse events will not be imputed. For evaluations other than adverse events and concomitant medications, missing or incomplete dates that are needed for analysis will also be imputed. If the day is missing for a study procedure or assessment but the month and year are present, the day will be imputed as the 15th of the month. Month and year will not be imputed for study procedures or assessments. These same rules will be applied to missing or incomplete dates for relapses.

## 6.2 Analysis populations, Treatment groupings, and Datasets

### 6.2.1 Analysis populations

The analysis populations of Study 109MS303 that will be utilized are described below.

- Study 109MS303 Intent-to-Treat (ITT) Population: This analysis population is defined to include all subjects who entered Study 109MS303 and received at least one dose of study treatment in Study 109MS303. Subjects will be analyzed in the treatment groups to which they were randomized/assigned. This will be the primary analysis population utilized for the majority of baseline analyses and non-MRI analyses addressing long-term efficacy and persistence of efficacy.

- Study 109MS303 MRI Cohort: This analysis population is defined to include subjects in the Study 109MS303 ITT Population who consented to participate in the MRI substudy and have any MRI data in Studies 109MS301, 109MS302, or 109MS303. This will be the primary analysis population utilized for the majority of MRI analyses addressing long-term efficacy and persistence of efficacy.

- Study 109MS303 Safety Population: This analysis population is defined to include all subjects who had any post-baseline safety follow-up in Study 109MS303, defined

as any treatment emergent AE in Study 109MS303 or any post-baseline laboratory, vital signs, or physical exam assessment in Study 109MS303, and received at least one dose of study treatment in Study 109MS303. Subjects will be analyzed in the treatment groups to which they were randomized/assigned.

### 6.2.2 Treatment Groupings

The analysis populations described above in Section 6.2.1 will be utilized in conjunction with the following three sets of treatment groupings as appropriate:

- Combined Treatment Grouping, indicating the study treatments during Study 109MS301 or 109MS302 and Study 109MS303 combined:
    - BG00012 240 mg BID -> 240 mg BID (i.e., BG00012 240 mg BID in both studies), or, alternatively, BG00012 240 mg BID Previously Treated with BG00012 240 mg BID;
    - BG00012 240 mg TID -> 240 mg TID (i.e., BG00012 240 mg TID in both studies), or, alternatively, BG00012 240 mg TID Previously Treated with BG00012 240 mg TID;
    - Placebo -> BG00012 240 mg BID (i.e., placebo in Studies 109MS301/302 and BG00012 240 mg BID in Study 109MS303), or, alternatively, BG00012 240 mg BID Previously Treated with Placebo;
    - Placebo -> BG00012 240 mg TID (i.e., placebo in Studies 109MS301/302 and BG00012 240 mg TID in Study 109MS303), or, alternatively, BG00012 240 mg TID Previously Treated with Placebo;
    - GA -> BG00012 240 mg BID (i.e., GA in Study 109MS302 and BG00012 240 mg BID in Study 109MS303), or, alternatively, BG00012 240 mg BID Previously Treated with GA;
    - GA -> BG00012 240 mg TID (i.e., GA in Study 109MS302 and BG00012 240 mg TID in Study 109MS303), or, alternatively, BG00012 240 mg TID Previously Treated with GA.

  Analyses utilizing the Study 109MS303 ITT Population, Study 109MS303 MRI Cohort, or Study 109MS303 Safety Population may employ Combined Treatment Grouping. The first treatment group name in each bullet above (e.g., BG00012 240 mg BID -> 240 mg BID) will be utilized for relevant efficacy summaries, while the second treatment group name in each bullet above (e.g., BG00012 240 mg BID Previously Treated with BG00012 240 mg BID) may be utilized for relevant safety summaries.

- Study 109MS303 Treatment Grouping, indicating the study treatment during Study 109MS303:
    - BG00012 240 mg BID;

- BG00012 240 mg TID; and
- Total BG00012 (i.e., BG00012 240 mg BID and BG00012 240 mg TID groups combined).

Analyses utilizing the Study 109MS303 Safety Population may employ Study 109MS303 Treatment Grouping. This treatment grouping will be utilized for select safety analyses where the incidence of an event or abnormality in Study 109MS303 itself may be of interest.

Baseline summaries will include both Combined Treatment Grouping and Study 109MS303 Treatment Grouping.

### 6.2.3 Datasets

The analysis populations and treatment groupings described above in Sections 6.2.1 and 6.2.2, respectively, will be utilized in conjunction with the following three datasets as appropriate:

- Combined Data, defined as data from Studies 109MS301, 109MS302, and 109MS303; and
- Study 109MS303 data only, defined as data from Study 109MS303.

Analyses utilizing Combined Treatment Grouping may include either Combined Data or Study 109MS303 data only. Analyses utilizing Study 109MS303 Treatment Grouping will include Study 109MS303 data only.

### 6.2.4 Summary of Use of Analysis populations, Treatment groupings, and Datasets

The following table summarizes the use of analysis populations by treatment groupings and datasets, with a brief mention of the planned analyses in italicized text in brackets:

**Table 6.2.4.1: Summary of use of analysis populations by treatment groupings and datasets**

| Dataset | Treatment Grouping | |
|---|---|---|
| | **Combined Treatment Grouping** | **Study 109MS303 Treatment Grouping** |
| **Combined Data** | Study 109MS303 ITT Population *[Majority of Non-MRI Efficacy Analyses]*<br><br>Study 109MS303 MRI Cohort *[Majority of MRI Efficacy Analyses]* | N/A |
| **Study 109MS303 data only** | Study 109MS303 ITT Population *[Study 109MS303 ARR Analysis]*<br><br>Study 109MS303 Safety Population *[Majority of Safety Analyses]* | Study 109MS303 Safety Population *[Selected Safety Analyses]* |

Notes: Italicized text in brackets provides a brief description of the analyses planned for the particular combination of analysis population, treatment grouping, and dataset. Lymphocyte analysis may be utilized with Combined Data and Combined Treatment Grouping or Study 109MS303 Treatment Grouping. N/A = Not Applicable.

## 6.3 Description of Analytic Methods

### 6.3.1 Analysis of Baseline Data

Baseline data will be summarized utilizing the Study 109MS303 ITT Population and both Combined Treatment Grouping and Study 109MS303 Treatment Grouping. The definition of baseline will be dependent on the analysis performed and the study data utilized.

The following are the various definitions of baseline that may be utilized in analyses:

- Study 109MS303 Baseline is defined as the latest available value prior to or on the date of first dose of study treatment in Study 109MS303;
- Previous Study Baseline is defined as the latest available value prior to or on the date of first dose of study treatment in Study 109MS301 or 109MS302; and
- BG00012 Treatment Baseline is defined as the latest available value prior to or on the date of first exposure to BG00012, specifically:

- the latest available value prior to or on the date of first exposure to study treatment in Study 109MS301 or 109MS302 for subjects on BG00012 treatment in Study 109MS301 or 109MS302;
- the latest available value prior to dosing in Study 109MS303 for subjects in the placebo or GA treatment groups in Study 109MS301 or 109MS302.

### 6.3.2 Accounting of Subjects

The following categories of accounting will be summarized for Study 109MS303 using numbers and percentages:

- Subjects who entered the study;
- Subjects who were dosed;
- Subjects who discontinued treatment (and therefore withdrew from study), along with reasons for discontinuing treatment; and
- Subjects who have lymphocyte follow up.

Listings of subjects who discontinued treatment and withdrew from the study and the reasons for discontinuation/withdrawal will be presented.

### 6.3.3 Protocol Deviations

Protocol deviations identified in the study will be listed.

### 6.3.4 Demographic and Baseline Characteristics

Demographic characteristics, including age (years), age category (18-19, 20-29, 30-39, 40-49, 50-55, and >55, as well as <40, ≥40 ), gender, and race, will be summarized. In addition, region, height (cm), weight (kg), and body mass index (kg/m^2) will also be summarized. Weight, height, and body mass index will be presented based on the Study 109MS303 Baseline definition, while age will be presented based on both the Previous Study Baseline (corresponding to the age covariate used in efficacy models) and Study 109MS303 Baseline definitions.

Baseline disease characteristics will also be summarized, including MS disease history (time since first MS symptoms and time since diagnosis of MS at Study 109MS303 Baseline), relapse history (number of relapses in the past 12 months and in the past 3 years at Previous Study Baseline), McDonald criteria (at Previous Study Baseline), and EDSS score (at both the Previous Study Baseline and Study 109MS303 Baseline). In addition, summary statistics for baseline MRI parameters(number and volume of Gd-enhancing lesions, number and volume of T2 lesions, number and volume of T1 lesions, normalized whole brain volume, and median MTR of whole brain) for the Study 109MS303 MRI cohort will also be presented, as applicable, at both the Previous Study Baseline and Study 109MS303 Baseline.

Prior MS treatment history and baseline cardiac risk factors will also be summarized based on the Previous Study Baseline definition.

### 6.3.5 Medical History

An updated medical history is collected at the baseline visit of Study 109MS303. A summary of this updated medical history will be presented.

### 6.3.6 Study Drug Compliance and Time on Study

Study drug compliance in Study 109MS303 will be summarized. The number of capsules not taken since the previous visit (out of those capsules the subject was expected to take) is collected on the study drug accountability CRF page. Using this information, percentage compliance will also be calculated as follows:

> *Total number of capsules not taken* is calculated as the sum of all individual values of the number of capsules not taken since the previous visit for a subject, as defined above.
>
> *Total number of capsules assigned* is calculated based on the number of days on study medication for a subject. The calculation of number of days on study medication is detailed below.
>
> *Percentage compliance* is then calculated as:
> [(*Total number of capsules assigned*) – (*Total number of capsules not taken*)] / (*Total number of capsules assigned*) * 100.

Percentage compliance will be calculated up to the date of last dose in Study 109MS303.

In addition, the total time on study medication will be summarized as a continuous variable and categorized into intervals. Days on study medication is calculated as the number of days from date of first dose to date of last dose. Additionally, the total number of subject-years of exposure to study medication will be presented. Time on study medication will be summarized for Study 109MS303 only and for Combined Data, utilizing the Study 109MS303 ITT and Safety populations.

Time on study parameters, including overall time on study and time on study prior to first use of alternative MS medication, will be summarized as a continuous variable and categorized into intervals. Total number of subject-years for time on study, both overall and prior to first use of alternative MS medication, will also be presented with respect to the dataset analyzed. The last known date on study will be taken as the last visit/evaluation date from relevant data for a subject. Summaries will be performed for Study 109MS303 only and for Combined Data, utilizing the Study 109MS303 ITT population. Additionally, overall time on study in Study 109MS303 will be summarized utilizing the Study 109MS303 Safety population.

### 6.3.7 Concomitant Medications and Non-drug Therapies

Concomitant medications are defined as any medications take on or after the day of first dose of study treatment and on or before the day of last does of study treatment in the given

dataset. This includes medications started prior to the first dose of study treatment in the given dataset and continued while on study or started after the first dose of study treatment. A windowing rule of +180 days may be applied to these subjects who have completed the study without lymphocyte follow-up. The WHO drug dictionary will be used for coding concomitant medications.

Concomitant medications will be summarized using the Study 109MS303 ITT Population and Study 109MS303 data only. The number and percent of subjects taking concomitant medications will be summarized. The most frequently taken concomitant medications, defined as those taken by at least 10% of subjects in any treatment group, will also be presented.

Additionally, summaries of first alternative MS medication used may be performed using the following analysis populations, treatment groupings, and datasets:

- Study 109MS303 ITT Population and Combined Data; and
- Study 109MS303 ITT Population and Study 109MS303 data only.

The number and percent of subjects in Study 109MS303 who previously took alternative MS medications in Studies 109MS301 or 109MS302 will also be summarized.

### 6.3.8 Day Ranges for Analyses by Visit

Subjects do not always strictly adhere to the visit schedule timing in protocols. Therefore, the designation of visits (or timepoints) will generally be based on the day of evaluation relative to date of first dose of study medication within a given dataset (i.e., with the date of first dose of study medication being Study Day 1), rather than the nominal visit recorded in the clinical database.

Mutually exclusive visit windows will generally be utilized to assign visits for analysis that correspond to post-baseline visits specified in the protocols. In general, for data that are summarized by visit, visits for analysis will be assigned by using a windowing scheme as described below.

Within a given dataset, the upper bound of the baseline visit window is Day 1, and the lower bound of the first post-baseline visit window is Day 2. For all other lower and upper bounds of visits, windows will end at the midpoint between scheduled visit timepoints, with the midpoint in the latter visit window.

If more than one observation occurs within the same window, data closest to the target day for the scheduled visit will be used for analysis. If more than one observation is the same distance from the target day, the last observation will be used for analysis.

However, for MRI analyses, visit windowing will only be applied when the nominal visit designation is premature withdrawal, with the lower bound of the first post-baseline visit window being Day 169 and with all other lower and upper visit bounds defined as above.

Otherwise, MRI analyses will not utilize visit windowing and will present results by nominal visit designations.

## 6.4 Efficacy Analysis

### 6.4.1 Analysis Population

Analysis populations, treatment groupings, and datasets are detailed in Section 6.2 above.

The analysis populations, treatment groupings, and datasets for efficacy analyses will be as follows:

- Primary efficacy analyses of non-MRI endpoints will utilize the Study 109MS303 ITT Population with Combined Treatment Grouping and Combined Data;
- Primary efficacy analyses of MRI endpoints will utilize the Study 109MS303 MRI Cohort with Combined Treatment Grouping and Combined Data;
- The efficacy analysis of Study 109MS303 ARR will utilize the Study 109MS303 ITT Population with Combined Treatment Grouping and Study 109MS303 data only.

### 6.4.2 Analysis Methods

In general, the covariates included in the statistical models for efficacy endpoints are selected to be consistent with analysis models in other previously reported Phase 3 MS studies. In these studies, the covariates have been demonstrated to be important factors in the analyses of the efficacy endpoints. Additionally, the statistical models and the covariates for inclusion in these statistical models for the efficacy analyses for Study 109MS303 will generally be the same as the corresponding models in Studies 109MS301 and 109MS302.

Consistent with the efficacy analysis conventions in Studies 109MS301 and 109MS302, in general, for efficacy analyses of Study 109MS303, observed data after first use of alternative MS medications in the given dataset will be excluded, or subjects will be censored at the time of first starting alternative MS medication in the given dataset if they have not experienced the event of interest. The exceptions to this are:

- Select sensitivity analyses that may be performed which consider all data regardless of alternative MS medication usage; and
- Analysis of disability progression based on EDSS, where EDSS evaluations performed after first use of alternative MS medication has been started will be used to confirm tentative progression that occurred prior to first use of alternative MS medication.

Alternative MS medications may include but are not limited to: interferon beta-1a, glatiramer acetate (for those subjects not receiving study-assigned GA), natalizumab, interferon beta-1b, mitoxantrone, mycophenolic acid, mycophenoloate mofetil, plasmapheresis used for the indication of MS, fingolimod, methotrexate, azathioprine,

alemetuzumab, ocrelizumab, teriflunomide, siponimod, daclizumab, cladribine, peginterferon beta-1a, cyclophosphamide, and rituximab.

Region as a covariate or in the subgroup analysis is based not only on considerations of geography but also on type of health care system and access to health care in each country. The categories for regions will be identical to those used in Studies 109MS301 and 109MS302:

- Region 1: United States (including Puerto Rico);
- Region 2: Western European countries (including Austria, Belgium, France, Germany, Greece, Ireland, Italy, Netherlands, Spain, Switzerland, and United Kingdom), plus Australia, Canada, Costa Rica, New Zealand, Israel, and South Africa; and
- Region 3: Eastern European countries (including Belarus, Bosnia And Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Latvia, Macedonia, Moldova, Poland, Romania, Serbia, Slovakia, and Ukraine), India, Guatemala, and Mexico.

The following table includes endpoints, analysis populations, baseline type with a short summary of statistical methods for the major efficacy analyses. Sections below include more details about these analyses.

#### 6.4.2.1 Table of summary of major efficacy analyses for the 109MS303 ITT population

| Endpoint | Population | Data | Treatment Grouping | Baseline | Statistical Methods |
|---|---|---|---|---|---|
| Annualized relapse rate | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Unadjusted and adjusted ARR. Adjusted ARR will be estimated from the negative binomial model, adjusted for baseline EDSS (<=2.0 vs >2.0), baseline age (<40 vs >=40), region and number of relapses in the 1 year prior to 301/302 study entry. If convergence criterion is not met, then Poisson regression will be used. |
| | | 109MS303 Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Unadjusted and adjusted ARR. Adjusted ARR will be estimated from the negative binomial model, adjusted for baseline EDSS (<=2.0 vs >2.0), baseline age (<40 vs >=40), region and number of relapses in the 1 year prior to 301/302 study entry. If convergence criterion is not met, then Poisson regression will be used. |
| Annualized relapse rate by yearly interval | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Unadjusted and adjusted ARR. Adjusted ARR will be estimated from the negative binomial model, adjusted for baseline EDSS (<=2.0 vs >2.0), baseline age (<40 vs >=40), region and number of relapses in the 1 year prior to 301/302 study entry. If convergence criterion is not met, then Poisson regression will be used. |
| | | 109MS303 Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Unadjusted and adjusted ARR. Adjusted ARR will be estimated from the negative binomial model, adjusted for baseline EDSS (<=2.0 vs >2.0), baseline age (<40 vs >=40), region and number of relapses in the 1 year prior to 301/302 study entry. If convergence criterion is not met, then Poisson regression will be used. |

| Endpoint | Population | Data | Treatment Grouping | Baseline | Statistical Methods |
|---|---|---|---|---|---|
| Proportion of subjects who relapsed at selected time points | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Kaplan-Meier curve of time to first relapse will be used to estimate the proportion at selected time points. |
| Proportion of subjects with 6-month sustained disability progression measured by EDSS at selected time points | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Kaplan Meier analysis of time to the 6-month sustained disability progression will be used to estimate the proportion at selected time points. |
| | | 109MS303 Data | Combined Treatment Grouping | 109MS303 Baseline | Kaplan Meier analysis of time to the 6-month sustained disability progression will be used to estimate the proportion at selected time points. |
| EQ-5D: EQ5D VAS, EQ-5D index score and EQ-5D individual questions | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Descriptive statistics by combined treatment group and by study visit. |
| SF-36 scores | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Descriptive statistics by combined treatment group and by study visit. |
| Visual function test | 109MS303 ITT population | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Descriptive statistics by combined treatment group and by study visit. |
| MRI endpoints | 109MS303 MRI Cohort | Combined Data | Combined Treatment Grouping | 109MS301/109MS302 Baseline | Descriptive statistics by combined treatment group and by study visit.<br>For the total number of new or newly enlarging T2 lesions, the total number of new T1 lesions, and the total number of Gd-enhancing lesions: negative binomial models may be used, adjusted for baseline characteristics in previous studies. |

### 6.4.3 Analysis of Relapse

Definition of Relapse

In Study 109MS303, relapses are defined as new or recurrent neurologic symptoms not associated with fever or infection, lasting at least 24 hours, and accompanied by new objective neurological findings upon examination by the *treating (or examining) neurologist*. The subject must have objective signs on the *treating (or examining) neurologist's* examination confirming the event (i.e., an objective or protocol-defined relapse). New or recurrent neurologic symptoms that evolve gradually over several weeks to months are not to be considered an acute relapse. New or recurrent neurologic symptoms that occur less than 30 days following the onset of a protocol-defined relapse should be considered part of the same relapse. i.e., if 2 relapses have onset days that are ≤29 days of one another, they will be counted only as 1 relapse, and the onset date used in the analysis will be the onset date of the first relapse.

In Studies 109MS301 and 109MS302, relapses are defined as new or recurrent neurologic symptoms not associated with fever or infection, lasting at least 24 hours, and accompanied by new objective neurological findings upon examination by the *examining neurologist*. The subject must have objective signs on the *examining neurologist's* examination confirming the event (i.e., an objective or protocol-defined relapse), and additionally must then be reviewed and confirmed by the *Independent Neurology Evaluation Committee (INEC*) (INEC-confirmed relapses). New or recurrent neurologic symptoms that evolve gradually over several weeks to months were not to be considered an acute relapse. New or recurrent neurologic symptoms that occur less than 30 days following the onset of an INEC confirmed relapse should be considered part of the same relapse. i.e., if 2 relapses have onset days that are ≤29 days of one another, they will be counted only as 1 relapse, and the onset date used in the analysis will be the onset date of the first relapse.

Therefore, Studies 109MS301, 109MS302, and 109MS303 all utilize essentially the same definition for objective (i.e., protocol-defined) relapses, given that there is no Independent Neurology Evaluation Committee (INEC) in Study 109MS303. Therefore, analyses of relapse for the final analysis of Study 109MS303 will be based primarily on objective (i.e., protocol-defined) relapses.

Annualized Relapse Rate

Annualized relapse rate (ARR) will be summarized using the following analysis populations, treatment grouping, and datasets:

- Studies 109MS301/302/303 ARR, overall and over time, using the Study 303 ITT Population with Combined Treatment Grouping and Combined Data; and
- Study 109MS303 ARR using the Study 303 ITT Population with Combined Treatment Grouping and Study 109MS303 data only.

The number of subjects with a given number of relapses (categories of 0, 1, 2, 3, and 4 or more relapses), total number of relapses, and total number of subject-years of follow-up will be summarized by treatment group. The unadjusted ARR will be calculated for each treatment group as the total number of relapses divided by the total number of subject-years

of follow-up. Relapses and follow-up time that occur after first use of alternative MS medications are excluded from the calculations.

The adjusted ARR and associated 95% confidence interval will be calculated for each treatment group by using a negative binomial regression model for unadjusted ARR, adjusting for baseline EDSS score (≤2.0 vs. >2.0), baseline age (<40 vs. ≥40 years old), region, and baseline relapse rate (i.e., the number of relapses in the year prior to the date of first dose in the previous study) as covariates. The baseline definition for these covariates will be based on Previous Study Baseline.

The logarithmic transformation of the time on the study will be included in the model as the "offset" parameter. If the data are underdispersed, or if the negative binomial regression model does not converge, a Poisson regression model with the same covariates will be used instead of the negative binomial regression model. Dispersion will be evaluated from the Pearson Chi-Square statistic. If the ratio of the Pearson Chi-Square statistic to the degrees of freedom is ≤ 1 which indicates no overdispersion, then a Poisson regression model with adjustment for underdispersion will be used.

In addition, the subject relapse rate will be calculated as the number of relapses for an individual subject divided by follow-up time in years for that subject excluding follow-up time after first use of alternative MS medication. Subject relapse rate will be summarized descriptively by treatment group.

ARR analyses over time will be presented by various time intervals in order to elucidate onset and maintenance of effect over time. For example, time intervals one year in length may be considered.

For ARR analyses, data following first use of alternative MS medications in the given dataset used for analysis will be excluded. If a subject did not experience a relapse prior to first starting an alternative MS medication, the subject's follow-up time will be censored at the time the first alternative MS medication is started. Likewise, if a subject did not experience a relapse prior to study withdrawal, the subject's follow-up time will be censored on the last date followed. The start date for calculation of ARR will be date of first dose, and if date of first dose is incomplete, date of randomization or enrollment will be used.

Proportion of Subjects Who Relapsed

The proportion of subjects who relapsed will be summarized using the following analysis populations, treatment grouping, and datasets:

- Studies 109MS301/302/303 proportion relapsed using Study 303 ITT Population with Combined Treatment Grouping and Combined Data.

The proportion of subjects relapsed can be estimated as the probability of relapses from the Kaplan-Meier curve of the time to the first relapse (i.e., Kaplan-Meier product-limit estimator). If there are no early withdrawals, the Kaplan-Meier estimate of the proportion of subjects relapsed is the same as the observed proportion of subjects relapsed. If there are early withdrawals, the Kaplan-Meier estimate has the advantage of taking into account the length of follow-up for the early withdrawals, without making any assumptions about whether those subjects relapsed or not.

Proportion relapsed analyses over time will be presented based on Kaplan-Meier curves. The estimated proportion of subjects relapsed at key timepoints from the Kaplan-Meier curves will be presented.

For proportion relapsed analyses, data following first use of alternative MS medications will be excluded. If a subject did not experience a relapse prior to first starting an alternative MS medication, the subject will be censored at the time the first alternative MS medication is started. Likewise, if a subject did not experience a relapse prior to study withdrawal, the subject will be censored on the last date followed. The start date for calculation of time to relapse will be date of first dose, and if date of first dose is incomplete, date of randomization will be used.

### 6.4.4 Analysis of Progression

Definition of Progression

In Study 109MS303, sustained disability progression will be assessed from the EDSS scores performed as per the protocol defined schedule of assessments at regular visits relative to the baseline EDSS score in the extension study, and well as from the baseline scores in previous study. Sustained disability progression is defined as at least a 1.0 point increase on the EDSS from a baseline EDSS ≥1.0 that is sustained for at least 24 weeks or a 1.5 point increase on the EDSS from a baseline EDSS = 0 that is sustained for at least 24 weeks. The EDSS progression is defined as sustained (or confirmed) when this minimum EDSS change is present on the next study visit occurring after 154 days or longer from the initial observation. The 154 day interval is based on the visit windows allowed in the protocol around the target visit day

The date of the initial visit at which the minimum increase in the EDSS is met will be the date of onset of the progression (tentative progression).

Progression will not be confirmed at a visit where a relapse is also occurring. A subject is considered to be having a relapse for at least 29 days after the start date of an objective relapse. If a subject meets the defined EDSS criteria of sustained progression and is also having a relapse, the subject will be required to meet the defined minimum EDSS criteria at the subsequent visit.

Disability progression can be confirmed at the early (premature) study withdrawal visit, according to the rules above, as long as the early withdrawal visit is not also a relapse assessment visit.

Death due to MS will be counted as progression. If the subject was in the midst of a tentative progression at the time of death (e.g. the EDSS evaluation prior to death is a tentative progression), the progression date will be the date of the start of the progression. If the subject had a confirmed EDSS progression prior to death, the progression date is the start date of the tentative progression. Otherwise, the progression date will be the date of death.

In Studies 109MS301 and 109MS302, the same definition of sustained disability progression will be used as in Study 109MS303 (i.e., requiring a 24 week confirmation).

Sustained disability progression will be summarized using the following analysis populations, treatment groupings, and datasets:

- Studies 109MS301/302/303 time to sustained (24 week) disability progression using the Study 303 ITT Population with Combined Treatment Grouping and Combined Data with previous study baseline; and

- Studies 109MS301/302/303 time to sustained (24 week) disability progression using the Study 303 ITT Population with Combined Treatment Grouping and Study 109MS303 data only with 109MS303 baseline.

Time to sustained disability progression will be presented based on Kaplan-Meier curves. The estimated proportion of subjects with sustained disability progression at key timepoints based on the Kaplan-Meier curve will be presented.

For these sustained disability progression analyses, subjects who do not have a sustained progression based on the above definition will be censored. The censor date will be the date of the last EDSS assessment, or the last EDSS assessment prior to first use of an alternative MS medication, unless the subject has a tentative progression at this assessment. For subjects with a tentative progression at this assessment, (or at the last EDSS assessment prior to first use of alternative MS medication that was subsequently not confirmed) the censor date will be the date of the EDSS assessment prior to the last EDSS assessment. Subjects who withdraw from the study after the baseline visit but prior to the first scheduled EDSS assessment will be censored at baseline. The start date for calculation of time to progression will be date of first dose, and if date of first dose is incomplete, date of randomization will be used.

### 6.4.5 Analysis of MRI Endpoints

The following MRI endpoints will be analyzed:

- Lesion parameters:
    - number of Gd-enhancing lesions;
    - number of new or newly-enlarging T2 lesions; and
    - number of new T1 hypointense lesions.

- Volume parameters:
    - volume of Gd-enhancing lesions;
    - volume of T2 hyperintense lesions; and
    - volume of T1 hypointense lesions.

- Other parameters:
    - brain atrophy (percent brain volume change [PBVC]); and
    - magnetization transfer ratio (MTR).

MRI endpoints will be summarized using the following analysis populations, treatment grouping, and datasets:

- Study 109MS303 MRI Cohort with Combined Treatment Grouping and Combined Data

The number of Gd-enhancing lesions will be descriptively summarized, both overall and over time, by treatment group, both as a continuous variable and categorically.

The number of new or newly-enlarging T2 lesions will be descriptively summarized, both overall and by key time intervals, by treatment group, both as a continuous variable and categorically. The negative binomial regression model will be used to estimate adjusted means and 95% confidence intervals for each treatment group. The model will adjust for region and baseline volume based on the Previous Study Baseline.

Examples of key time intervals may include, but are not limited to: Previous Study Baseline to Week 24 in Previous Study, Week 24 to Week 48 in Previous Study, Week 48 to Week 96 in Previous Study, Week 96 of Previous Study to Week 48 of Study 109MS303, Week 48 of Study 109MS303 to Week 96 of Study 109MS303, and up to Week 336 of Study 109MS303 to Week 384 of Study 109MS303, Previous Study Baseline to Week 48 of Previous Study, Previous Study Baseline to Week 96 of Previous Study, Previous Study Baseline to Week 48 of Study 109MS303, and up to Previous Study Baseline to Week 384.

The number of new T1 hypointense lesions will be analyzed in a similar manner as the number of new or newly-enlarging T2 lesions.

The actual value of MRI volume parameters, as well as the change and percentage change in these parameters for key time intervals (similar to those used for analysis of T2 lesions), will be descriptively summarized by treatment group.

Brain atrophy (PBVC) and MTR will also be descriptively summarized by treatment group for key timepoints and time intervals.

Analyses utilizing the Studies 109MS301 + 109MS302 MRI Cohort will descriptively present the number of new lesions, as well as change in lesion volume for key time intervals (e.g., Previous Study Baseline to Week 48 in Previous Study, Week 48 to Week 96 in Previous Study, Week 96 in Previous Study to Week 48 in Study 109MS303, Week 48 in Study 109MS303 to Week 96 in Study 109MS303, Week 48 in Study 109MS303 to Week 96 in Study 109MS303, Week 96 in Study 109MS303 to Week 144 in Study 109MS303, Week 144 in Study 109MS303 to Week 192 in Study 109MS303, and up to Week 336 in Study 109MS303 to Week 384 in Study 109MS303).

As a result of Study 109MS303 becoming an open label extension with one single arm after regulatory approved the dosage of BG00012 240 mg BID, all patients receiving BG00012 240 mg TID were switched to BG00012 240 mg BID. Hypothesis tests of comparison between treatment groups for MRI endpoints will not be conducted.

For MRI analyses, data following first use of alternative MS medications will be excluded. Some MRI endpoints for Study 109MS303 MRI Cohort with Combined Treatment Grouping and 109MS303 Data may be analyzed.

### 6.4.6 Analysis Quality of life Assessments (SF-36 and EQ-5D) and Visual Function Test

The quality of life (SF-36 and EQ-5D) and Visual Function Test assessments will be summarized using the Study 109MS303 ITT Population with Combined Treatment Grouping and Combined Data. For these analyses, data following first use of alternative MS medications will be excluded.

**SF-36**

Scores for SF-36 will be calculated based on the standard scoring method described in the SF-36 Health Survey Manual and Interpretation Guide. This method indicates that for an individual domain where 50% or more of the item's questions are non-missing, the average of the non-missing items for that patient in that domain will be calculated, and this average is then used to impute the missing item questions response and calculate the domain score. The mental and physical component summary measures (MCS and PCS, respectively) are then calculated from the imputed domains scores.

Actual scores and change from Previous Study Baseline over time will be descriptively summarized by treatment group for the individual domains and MCS and PCS measures.

**EQ-5D**

The Euroqol (EQ)-5D is a generic health-related quality of life instrument which has been extensively validated. There are two components to the EQ-5D, a Health State Profile that results in calculation of an EQ-5D index score and a Visual Analog Scale (VAS) that results in calculation of the subject assessment of health scale.

For the Health State Profile, patients record their level of current health for five domains comprising a health profile: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Scores of 1, 2, or 3 are possible responses for each of the five questions in the Health State Profile (1=no problems, 2=some problems, 3=severe problems). The digits for the five dimensions are combined in a 5-digit number describing a person's health state. A scoring formula developed by the EuroQol Group is then used to assign utility values for each patient's Health State Profile. A summary index score (EQ-5D index score) is derived from the 5 questions by conversion with this scoring formula and a table of scores.

For the VAS, patients are asked to rate their current health on a 20 cm scale from 0 to 100 where 0 represents "worst imaginable health state" and 100 represents "best imaginable health state" resulting in the subject assessment of health scale.

Actual scores and change from Previous Study Baseline over time will be descriptively summarized by treatment group for both the EQ-5D index score and subject assessment of health scale. Additionally, the scores for each of the five component questions of the EQ-5D index score will be summarized categorically.

**Visual Function Test**

The visual function test uses the low-contrast Sloan letter chart as an indicator of disease-related changes in visual performance. Subjects are tested using the contrast level of 100%, 2.5% and 1.25% charts, and the scores are defined as the number of letters identified correctly for each chart (maximum score is 60).

Actual scores and change from Previous Study Baseline over time will be descriptively summarized by treatment group for the VFT for each of the contrast levels.

## 6.5 Safety Data

### 6.5.1 Analysis Population

Analysis populations, treatment groupings, and datasets are detailed in Section 6.2 above.

Study 109MS303 Safety Population is defined to include all subjects who received at least one dose of study treatment in Study 109MS303 and had any post-baseline safety follow-up in Study 109MS303, defined as any treatment emergent AE in Study 109MS303 or any post-baseline laboratory, vital signs, or physical exam assessment in Study 109MS303. This definition of the safety population is utilized in order not to inflate the denominator for calculation of incidences and percentages (and therefore deflate the incidences and percentages themselves) with subjects who were dosed in Study 109MS303 but had no post-baseline safety follow-up.

The analysis populations, treatment groupings, and datasets for safety analyses will be as follows:

- Primary safety analyses will utilize the Study 109MS303 Safety Population with Combined Treatment Grouping and Study 109MS303 data only; and

- Select safety analyses (where the incidence in Study 109MS303 itself may be of interest) will utilize the Study 109MS303 Safety Population with Study 109MS303 Treatment Grouping and Study 109MS303 data only.

### 6.5.2 Analysis Methods

Safety analyses will summarize adverse events (AEs), laboratory abnormalities, and vital signs.

### 6.5.3 Analysis of Adverse Events

All summary analyses of AEs will be based on the principle of treatment emergence. In a given dataset, an event is considered to be treatment emergent if it has an onset date on or after the date of first study treatment or if it was present prior to start of study treatment and subsequently worsened. Henceforth, whenever an analysis or summary of AEs is mentioned, it is intended that this is in reference to the treatment emergent adverse event experience, even if not explicitly mentioned.

In general, analyses of AEs will include all events through the end of a subject's follow-up in a given dataset.

In general, AEs will be analyzed based on *incidence*, defined as the proportion of subjects who had at least one occurrence of an event out of the number of subjects in the relevant safety analysis population.

The Medical Dictionary for Regulatory Activities (MedDRA) version 13.1 will be utilized to code and group AEs by System Organ Class (SOC) and Preferred Term (PT). References to summaries of AEs using System Organ Class are intended to refer to the primary SOC. In addition to summaries of incidence by both SOC and PT, incidence of adverse events may also be summarized and presented by SOC only or by PT only, as appropriate, depending on the nature of the analysis performed. Adverse event summaries may also be provided by

High-Level Group Term (HLGT) or High-Level Term (HLT), if considered appropriate for signal detection, in instances where the SOC level of summary is too broad but the PT level of summary is too granular.

A summary analysis of adverse events will be provided including the following numbers and incidences/percentages of subjects: in the safety analysis population; with an adverse event; with a moderate or severe AE; with a severe AE; with a related AE; with a serious adverse event (SAE); with an AE leading to discontinuation of study treatment; and with an AE leading to withdrawal from study.

Summaries of the most frequently reported AEs by PT will also be presented, defined as those AEs with an incidence of AEs of at least 5% (at SOC or PT level) in any treatment group. Other cutoff values may be explored, as appropriate.

The incidence of AEs by SOC and PT may also be presented by time intervals (e.g., using a 12 month time interval). Other time intervals may be explored as well to elucidate trends over time. For such analyses, for a given time interval, the number of subjects who were followed for adverse events during that time interval will be presented along with the AE incidence during that time interval. Therefore, for a given SOC or PT, subjects will be counted only once for a given time interval but may be counted more than once across time intervals.

The incidence of AEs by relationship to study treatment will be summarized by SOC and PT. For a given SOC or PT, subjects with multiple AEs differing in relatedness will be counted once in the most related category. In addition, AEs categorized as “related” (i.e., those where the investigator assessment of relationship to study treatment is “possibly” or “related) will also be summarized by PT.

The incidence of AEs by severity will be summarized by SOC and PT using severity categories of mild, moderate, and severe. For a given SOC or PT, subjects with multiple AEs differing in severity will be counted once in the category of worst severity. The incidence of severe AEs will also be summarized by PT.

Deaths that occur will be described in detail in patient narratives. A listing of deaths will also be provided. While the number of deaths is expected to be small, tabulation will be performed if appropriate.

A summary of SAEs by SOC and PT will also be provided. Details of each SAE will also be listed, including subject ID, treatment group, system organ class/preferred term, date of onset, severity, and relationship to the study drug and action taken. In addition, a summary of treatment-related SAEs by SOC and PT will be provided.

The incidence of adverse events that led to study treatment discontinuation, study withdrawal, dose interruption, and dose reduction will each be presented by SOC and PT.

A listing of AEs leading to study treatment discontinuation and study withdrawal will also be presented. Adverse events that were treatment-emergent to the previous study yet led to study treatment discontinuation and study withdrawal in Study 109MS303 will be denoted by a negative onset day in the listing.

The incidence of adverse events of special interest will be summarized by SOC and PT. The adverse event of special interest categories may include the following:

- Flushing and other related symptoms;
- Gastrointestinal tolerability;
- Infections, including potential opportunistic infections;
- Ischaemic cardiovascular disorders;
- Hepatic disorders;
- Renal disorders;
- Malignancies;
- Lymphopenia and Leukopenia.

Subjects who discontinue study treatment are to have a premature study withdrawal visit within four weeks of the last dose of study treatment dose and a telephone follow-up 12 weeks after the last dose of study treatment. For subjects who have discontinued study treatment, a separate summary of adverse events with onset after discontinuation of study treatment may be presented.

Subject narratives will be provided for the following event types in Study 109MS303:

- Deaths;
- Serious adverse events (only selected SAEs may be included);
- AEs leading to study drug discontinuation;
- AEs leading to study withdrawal;
- Special interest AE categories of potential opportunistic infections, ischaemic cardiovascular disorders, hepatic disorders, and malignancies;
- ALT or AST values >5xULN at any time post-baseline; and
- ALT or AST values >=3xULN and concurrently total bilirubin values > 2xULN.

Select AE analyses utilizing the Study 109MS303 Safety Population with Study 109MS303 Treatment Grouping and Study 109MS303 data only, mentioned in the second bullet in Section 6.5.1 above, may include, but are not limited to: summary analysis of adverse events (AEs), incidence of AEs by SOC and PT, incidence of most common AEs, incidence of treatment-related AEs, incidence of severe AEs, incidence of serious adverse events (SAEs), incidence of treatment-related SAEs, incidence of AEs that led to discontinuation of study drug, incidence of AEs that led to study withdrawal, and incidence of AEs of special interest such as malignancies, hepatic disorders, GI, opportunistic infections, herpes zoster, etc.

### 6.5.4 Analysis of Clinical Laboratory Data

The following clinical laboratory parameters are assessed in the protocol:

- Hematology: hemoglobin, hematocrit, red blood cell count, WBC count (with differential) and platelet count.
- Blood chemistry: sodium, potassium, chloride, total bilirubin, alkaline phosphatase, alanine transaminase/serum glutamate-pyruvate transaminase (ALT/SGPT), aspartate transaminase/serum glutamic-oxaloacetic transaminase (AST/SGOT), lactate dehydrogenase, gamma-glutamyl-transferase (GGT), blood urea nitrogen (BUN), creatinine, bicarbonate, calcium, magnesium, phosphate, uric acid, and glucose.
- Urinalysis: color, specific gravity, pH, protein, glucose, blood, ketones, microscopy, $\beta_2$-microglobulin, and microalbumin.
- Urine cytology

Analyses for laboratory data involving baseline value will be based on BG00012 Treatment Baseline (as defined above in Section 6.3.1).

*Shift analyses*

Laboratory data will be summarized using shift tables where appropriate. Each subject's hematology and blood chemistry values will be flagged as "low", "normal", or "high" relative to the normal ranges of the central laboratory, or as "unknown" if no result is available. Each subject's urinalysis values will be flagged as "positive", "negative", or if no value is available, "unknown".

Shifts from baseline to high/low status for hematology and blood chemistry parameters and shifts from baseline to high/positive status for urinalysis will be presented. Subjects need to have at least one post-baseline evaluation in order to be included in the analysis. In addition, shifts from baseline to the worst post-baseline value will be presented for relevant laboratory tests by treatment group by clinically relevant categories (e.g., <= ULN, >1 - <3 xULN, >=3 - 5 xULN, >5 – 10 xULN, >10 – 20 x ULN, >20 x ULN, etc.).

For urine cytology tests results, the number and percentage of subjects with an abnormal post-baseline result will be summarized by treatment group.

*Summary of laboratory values over time and other analyses*

In addition to the shift analyses, summary of actual values, change from baseline values, and percentage change from baseline values will be presented by visit for quantitative laboratory parameters. Graphs showing mean actual values or mean change from baseline values over time may be presented as well.

In addition, summaries of worst post-baseline laboratory values by clinically relevant categories may also be presented for selected parameters of interest by treatment group. The data may also be presented in graphs. For example, for liver function tests (alkaline phosphatase, ALT, AST, GGT and total bilirubin), categories may be defined based on cutoff values relative to the upper limit of normal (e.g., <= ULN, > ULN, >=3 x ULN, >5 x ULN, >10 xULN, and >20 xULN for ALT, AST, and GGT, and <= ULN, > ULN, > 1.5

xULN, and >2 xULN for total bilirubin). For hematology parameters, categories may be defined based on potentially clinically significant cutoffs (e.g., < 3.0 x 10^9/L and >= 16 x 10^9/L for WBC and <0.8 x 10^9/L, <0.5 x 10^9/L, and > 12 x 10^9/L for lymphocytes). For qualitative urinalysis parameters, categories may be defined based on the qualitative categories (e.g., normal/negative, trace, 1+, 2+, 3+, etc.).

Worst post-baseline laboratory value summaries will comprise those select safety analyses utilizing the Study 109MS303 Safety Population with Study 109MS303 Treatment Grouping and Study 109MS303 data only, mentioned in the second bullet in Section 6.5.1 above.

### 6.5.5 Analysis of Vital Signs Data

The analysis of vital signs will focus on the incidence of clinically relevant abnormalities. Analyses for vital signs data involving baseline value will be based on BG00012 Treatment Baseline (as defined above in Section 6.3.1).

The number of subjects evaluated and the number and percentage of subjects with clinically relevant post-baseline abnormalities will be presented by treatment group. The criteria for clinically relevant post-baseline abnormalities (from Table 2 of the protocol) are shown below in Table 6.5.5.1 . Summary statistics for actual values and change from baseline will also be presented.

**Table 6.5.5.1 Criteria to Determine Clinically Relevant Abnormalities in Vital Signs**

| Vital Sign | Criteria for Abnormalities |
|---|---|
| Temperature | >38ºC and an increase from pre-dosing of at least 1ºC |
| Pulse | >120 beats per minute post-baseline, or |
| | an increase from pre-dosing of more than 20 beats per minute, or |
| | <50 beats per minute post-baseline, or |
| | a decrease from pre-dosing of more than 20 beats per minute |
| Systolic Blood Pressure | >180 mmHg post-baseline, or |
| | an increase from pre-dosing of more than 40 mmHg, or |
| | <90 mmHg post-baseline, or |
| | a decrease from pre-dosing of more than 30 mmHg |
| Diastolic Blood Pressure | >105 mmHg post-baseline, or |
| | an increase from pre-dosing of more than 30 mmHg, or |
| | <50 mmHg post-baseline, or |
| | a decrease from pre-dosing of more than 20 mmHg |

Summaries of actual value and change from baseline (based on BG00012 Treatment Baseline) by visit for body weight will also be provided.

### 6.5.6 Analysis of Absolute Lymphocyte Counts

Subjects who temporarily withhold, or permanently discontinue study treatment for any reason and have a lymphocyte count <LLN will continue protocol-required visits and assessments and will also be followed up every 4 weeks for 24 weeks, then every 12 weeks (unless clinically indicated more often or at the Investigator's discretion) until the lymphocyte count is ≥LLN or for up to 48 weeks following drug discontinuation, whichever occurs sooner. [redacted]

In general, the lymphocyte follow-up analysis may include subjects who have at least one ALC < LLN during the treatment period or at the treatment discontinuation and have at least one ALC value during the lymphocyte follow-up period. Patients who have any post lymphocyte follow-up baseline values may be included in the follow-up analysis. ALCs collected during the lymphocyte follow-up period will be analyzed.

The lymphocyte follow-up period is defined as the time starting from the date of temporary dose interruption or study drug discontinuation. The start date of the lymphocyte follow-up period is defined as the start date of temporary dose interruption or the treatment end date. The last observed value collected on or prior to the date of temporary dose interruption or study drug discontinuation will be the reference point (lymphocyte follow-up baseline). Data that are collected during the lymphocyte follow-up period will be assigned to an appropriate visit by using the mid-point of the two visits windowing scheme. Visits will be mapped every 4 weeks for 24 weeks and then every 12 weeks afterward.

The by-visit summary of actual value change and percent change from BG00012 baseline and the lymphocyte follow-up baseline in ALC during the lymphocyte follow-up period will be presented. The by-visit summary of actual value change and percent change from the study baseline in ALC during the whole study period will also be presented. In addition, graphical display of median (Q1/Q3) of actual value in ALC in the whole study period will be presented; graphical display of median (Q1/Q3) of percent change from BG00012 baseline and lymphocyte follow-up baseline will be presented. [redacted]

[redacted]

### 6.5.8 Deviations in Analyses from Study Protocol

With regulatory approval (effective March 2013), all patients receiving BG00012 240 mg TID switched to BID (approved dosage) at their next study visit. This study becomes an open label extension study with a single arm. Following analyses proposed in the protocol will not be conducted.

Cox PH analyses for time for first relapse and time to disability progression will not be conducted as we only have a single treatment arm.

ANCOVA models for MRI outcomes will not be conducted as we only have one treatment arm.