16.1.9 Documentation of Statistical Methods

The final approved Statistical Analysis Plans and other statistical documents, as applicable, for this study are provided in the following pages.

- Statistical Analysis Plans
    - Summary of Changes (all versions)
    - V4.0
    - V3.0
    - V2.0
    - V1.0
- Data Monitoring Committee Charter
- Data Monitoring Committee Decisions and Meeting Minutes
    - 22 Aug 2013 Recommendation
    - 19 Nov 2013 Recommendation
    - 25 Mar 2014 Recommendation
    - 17 Jun 2014 Recommendation
    - 04 Nov 2014 Recommendation
    - 19 Mar 2015 Recommendation
    - 29 Sep 2015 Recommendation
    - 24 Mar 2016 Recommendation
    - 20 Sep 2016 Recommendation
    - 27 Jan 2017 Recommendation
    - 20 Sep 2017 Meeting Minutes
    - 27 Mar 2018 Meeting Minutes

- Interim Analysis Reports
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 196 Subjects (Unblinded report and randomization ratios) – 03 Feb 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 250 Subjects (Unblinded report and randomization ratios) – 21 Apr 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 300 Subjects (Unblinded report and randomization ratios) – 16 Jun 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 350 Subjects (Unblinded report and randomization ratios) – 26 Sep 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 350 Subjects (Unblinded report and randomization ratios) – Addendum - 26 Sep 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 400 Subjects (Unblinded report and randomization ratios) – 19 Dec 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 400 Subjects (Unblinded report and randomization ratios) – Addendum – 19 Dec 2014
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 450 Subjects (Unblinded report and randomization ratios) – 14 Apr 2015
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 450 Subjects (Unblinded report and randomization ratios) – Addendum – 14 Apr 2015
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 500 Subjects (Unblinded report and randomization ratios) – 19 Jun 2015
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 500 Subjects (Unblinded report and randomization ratios) – Addendum – 19 Jun 2015
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 550 Subjects (Unblinded report and randomization ratios) – 07 Sep 2015
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 550 Subjects (Unblinded report and randomization ratios) – Addendum – 07 Sep 2015
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 600 Subjects (Unblinded report and randomization ratios) – 11 Jan 2016
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 600 Subjects (Unblinded report and randomization ratios) – Addendum – 11 Jan 2016
    - BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 650 Subjects (Unblinded report and randomization ratios) – 04 Apr 2016

- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 650 Subjects (Unblinded report and randomization ratios) – Addendum – 04 Apr 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 700 Subjects (Unblinded report and randomization ratios) – 09 Jun 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 700 Subjects (Unblinded report and randomization ratios) – Addendum – 09 Jun 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 750 Subjects (Unblinded report and randomization ratios) – 08 Aug 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 750 Subjects (Unblinded report and randomization ratios) – Addendum – 08 Aug 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 800 Subjects (Unblinded report and randomization ratios) – 29 Sep 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 800 Subjects (Unblinded report and randomization ratios) – Addendum – 29 Sep 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 3 Months post-accrual (Unblinded report and randomization ratios) – 21 Dec 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 3 Months post-accrual (Unblinded report and randomization ratios) – Addendum – 21 Dec 2016
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 6 Months post-accrual (Unblinded report and randomization ratios) – 17 Mar 2017
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 6 Months post-accrual (Unblinded report and randomization ratios) – Addendum – 17 Mar 2017
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 9 Months post-accrual (Unblinded report and randomization ratios) – 08 Jun 2017
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 9 Months post-accrual (Unblinded report and randomization ratios) – Addendum – 08 Jun 2017
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 12 Months post-accrual (Unblinded report and randomization ratios) – 05 Dec 2017
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis at 12 Months post-accrual (Unblinded report and randomization ratios) – Addendum – 05 Dec 2017

- BAN2401-G000-201 Tessella Interim Report – Interim Analysis (Sensitivity Analysis) at 12 Months post-accrual (Unblinded report and randomization ratios) – 05 Dec 2017
- BAN2401-G000-201 Tessella Interim Report – Interim Analysis (Sensitivity Analysis) at 12 Months post-accrual (Unblinded report and randomization ratios) – Addendum – 05 Dec 2017

REVISION HISTORY

Revisions to Version 2.0 (per Amendment 01)
Date: 10 Feb 2017

Change	Rationale	Affected Sections
Added exploratory objectives to evaluate Japanese subjects	To explore preliminary data in Japanese subjects for consistency of treatment effect between populations	• 3.1.3 • 3.2 • 5.3.1 • 5.6 • 5.6.2 • 5.8 • 6
Added exploratory objectives for pharmacokinetics (PK)	To characterize population PK and examine effect of intrinsic and extrinsic factors on the PK	• 3.1.3
Increased study duration; added potential for extension phase	To assess long-term safety and tolerability in subjects who elect to continue open-label treatment. Subjects will be administered 1 dose throughout the Open-label Extension Phase; 1 or 2 doses of BAN2401 will be studied in the Extension Phase, and dosing will be established upon early success.	• 3.2 • Figure 1 • 5
apolipoprotein E4 (*APOE4*) status must be confirmed before randomization and subjects who are confirmed *APOE4* positive not be randomized to the 10 mg/kg biweekly dose	The Voluntary Harmonisation Procedure (VHP) committee requested that subjects who are confirmed *APOE4* positive (*APOE4* hetero- or homozygous) not be randomized to the 10 mg/kg, biweekly dose.	• 3.2 • Table 1 • 5.2.4 • 5.3.2 • 5.3.4 • 5.6.2 • 6.3
Description of home infusion	Operational optimization of method of study drug administration due to low number of participation in home infusion option	• 3.2
Added definition of visit windows	To describe general rule to cover different visit schedules	• 8.1 • Table 2 (deleted)
Revised imaging substudy sample size from 260 to 306	To increase the statistical power and increase likelihood of detecting an effect on amyloid	• 4
Revised text regarding when statistical analyses will be done	Clarification	• 5

Revisions to Version 2.0 (per Amendment 01)
Date: 10 Feb 2017

Change	Rationale	Affected Sections
Clarification of measurement of secondary endpoint	Clarification; definition of amyloid positron emission tomography (PET) standard uptake value ratio (SUVR) added	• 5.1.2 • 5.4.4
Clarification of statistical method for vMRI and amyloid PET	Clarification	• 5 • 5.4.4
Clarifications to subject disposition, demographic and other baseline characteristics, prior and concomitant medications	Clarification	• 5.2.2 • 5.2.4 • 5.2.5
Clarification of adjustment for Covariates of Interest	Clarification	• 5.3.2
Clarification of subgroup analysis	Clarification	• 5.3.4
Clarification of approach to handling of missing data, sensitivity analyses, exploratory analyses	Clarification; The primary efficacy analysis is based on Bayesian statistics	• 5.3.5 • 5.4.1.1 • 5.4.1.2 • 5.4.3 • 5.8
Clarification of definition of treatment compliance	Clarification	• 5.2.6
Clarification of cerebrospinal fluid (CSF) biomarker analyses for values out of quantitation	Clarification	• 5.3.5
Clarification of statistical model using conventional analyses and timing, sensitivity analysis	Clarification	• 5.4.1.2
Clarification of report of Bayesian Aanalysis of primary endpoint in adaptive trial	Clarification	• 5.4.3
Clarification of extent of exposure analysis	Clarification	• 5.6.1
Clarification of definition of treatment-emergent adverse event (TEAE) and analyses	Clarification	• 5.6.2
Clarification of C-SSRS analysis	Clarification	• 5.6.3
Clarifications to anti-drug antibody (ADA) analysis	Clarification	• 5.6.4
Clarification of exploratory analyses for clinical endpoints and biomarkers	Clarification	• 5.8

Revisions to Version 2.0 (per Amendment 01)
Date: 10 Feb 2017

Change	Rationale	Affected Sections
Added study simulation plan	To clarify the performance of study design	• 4 • Appendix 13.2
Clarification of algorithms for efficacy parameters	Clarification; excluded not efficacy parameters	• 8.2
Update signature page	Personnel change	• Signature page
Grammatical, typographical, and formatting changes	Document quality and consistency	• Throughout

Revisions to Version 3.0 (per Amendment 02)
Date: 16 Nov 2017

Change	Rationale	Affected Sections
Updated key secondary objectives and corresponding endpoints and analyses	To establish the dose regimen with at least 90% of the maximum effective dose treatment effect (ED_{90}) dose of BAN2401 and to evaluate the effect of the ED_{90} dose of BAN2401 compared to placebo through conventional analysis; to evaluate potential disease-modifying effects of ED_{90} dose of BAN2401 compared to placebo on brain amyloid levels at 18 months	• 3.1.2 • 5.1.2 • 5.4.1.2 • 5.4.4
Updated secondary objectives and corresponding endpoint for evaluation of brain amyloid levels at 12 months	In alignment with key secondary objective update	• 3.1.3 • 5.1.3 • 5.4.4
Clarification of timing of 12-month Bayesian analysis and extension of the Bayesian analysis to 18 month endpoints	Clarification and in alignment with key secondary objective update	• 3.1.2 • 3.2 • 4 • 5 • 5.1.2 • 5.4.1 • 5.4.1.2 • 5.4.2 • 5.4.3 • 5.4.4 • 6 • 6.1

Revisions to Version 3.0 (per Amendment 02)
Date: 16 Nov 2017

Change	Rationale	Affected Sections
Updated justification for multiplicity adjustments for multiple dose groups in key secondary endpoints	For alignment with updates to key secondary objectives/endpoints	• 5.3.3
Clarified analysis of anti-drug antibodies during the study, including definitions for positive, negative, incidence rate, and placebo-treated subjects.	Clarification	• 5.6.4
Updated exploratory analyses for biomarkers	To describe the amyloid PET SUVR of global cortical average with whole cerebellum as reference region and subcortical white matter as longitudinal adjustment factor as exploratory analyses for biomarkers instead of secondary analyses	• 5.8
Grammatical, typographical, and formatting changes	Document quality and consistency	• Throughout

Revisions to Version 4.0 (per Amendment 03)
Date: 05 Jun 2018

Change	Rationale	Affected Sections
Removal of Extension Phase of the study	The Extension Phase of this study will not be conducted as the conditions defined in study protocol were not met	• 3 • 3.2 • Figure 1 • 5.2.6
Update of Key Secondary Objectives and associated endpoints and analyses	Rearranged key secondary objectives to emphasize disease pathophysiology based on 18 month data. Updated analysis methods to account for the lack of subjects positive for apolipoprotein ϵ4 variant (*APOE4*) in the 10 mg/kg biweekly dose group due to a change in the middle of randomization following a Regulatory request by European Health Authorities in July 2014.	• 3.1.2 • 5.1.2 • 5.3.3 • 5.3.4 • 5.3.5 • 5.4.4

Revisions to Version 4.0 (per Amendment 03)
Date: 05 Jun 2018

Change	Rationale	Affected Sections
Update of Secondary Objectives and associated endpoints and analyses	Rearranged secondary objectives to emphasize disease pathophysiology. Updated analysis methods to account for the lack of subjects positive for apolipoprotein ϵ4 variant (*APOE4*) in the 10 mg/kg biweekly dose group due to a change in the middle of randomization following a Regulatory request by European Health Authorities in July 2014.	• 3.1.3 • 5.1.3 • 5.3.4 • 5.3.5 • 5.4.4
Update of Exploratory Objectives and associated endpoints	In alignment with key secondary objective and secondary objective updates	• 3.1.4 • 5.1.4 • 5.3.4 • 5.3.5 • 5.8
Clarification of the definition of success at primary analysis and significance level to be used for all statistical tests for conventional analysis	Clarification	• 5
Clarification of presentation of demographic and baseline characteristics	Clarification	• 5.2.4
Clarification of the MMRM	Clarification	• 5.4.1.2 • 5.4.4
Added ANCOVA model with treatment as a factor and baseline value as a covariate	To evaluate treatment effect with simple model	• 5.4.1.2 • 5.4.4
Added the several Bayesian analysis	To evaluate treatment effect in the several ways	• 5.4.4
Clarification of PET analyses based on each tracer	Clarification; per small sample size with Flutemetamol	• 5.4.4
Update of Other Analyses section	To allow for inclusion of other analyses to be conducted as ad-hoc	• 5.7
Clarification of reference region in PET SUVR analyses	Clarification	• 5.8
Added ANCOVA for the relationship of change between clinical endpoints and biomarkers,	To evaluate treatment effect in the several ways	• 5.8

Revisions to Version 4.0 (per Amendment 03)
Date: 05 Jun 2018

Change	Rationale	Affected Sections
the proportion of subjects who have become amyloid negative, and Centiloid scale analysis		
Update to "Changes in the Planned Analyses" section	To reflect that the document has been updated to align with updates to the study protocol	• 7
Update of references section	Update to support secondary and exploratory analyses	• 12
Grammatical, typographical, and formatting changes	Document quality and consistency	• Throughout

STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BAN2401-G000-201
Study Protocol Title:	A Placebo-controlled, Double-blind, Parallel-group, Bayesian Adaptive Randomization Design and Dose Regimen-finding Study To Evaluate Safety, Tolerability and Efficacy of BAN2401 in Subjects With Early Alzheimer's Disease
Date:	v1.0, 22 Jan 2014 (Original SAP) v2.0, 10 Feb 2017 (Amendment 01) v3.0, 16 Nov 2017 (Amendment 02) v4.0, 05 Jun 2018 (Amendment 03)

SIGNATURE PAGE

Author:

PPD
Associate Director, Biostatistics
Neurology Business Group

[electronic signature in eDMS]
Signature

Date: _____

Approval

Biostatistics:
PPD Ph.D.
PPD Biostatistics and Project Operations (Head)
Neurology Business Group

[electronic signature in eDMS]
Signature

Date: _____

Biostatistics:
PPD Ph.D.
Senior Director, Biostatistics Project Lead
Neurology Business Group

[electronic signature in eDMS]
Signature

Date: _____

Study Director:
PPD PhD
Senior Director, Clinical Neuroscience
Neurology Business Group

[electronic signature in eDMS]
Signature

Date: _____

1 TABLE OF CONTENTS

2 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AD	Alzheimer's Disease
ADA	Ant-drug antibodies
ADAS-Cog	Alzheimer's Disease Assessment Scale – Cognitive Subscale
ADCOMS	Alzheimer's Disease Composite Score
AE	adverse event
AChEI	Acetylcholinesterase inhibitors
ANCOVA	analysis of covariance
ANOVA	analysis of variance
APOE4	Apolipoprotein ϵ4 variant
ARIA	amyloid related imaging abnormalities
ATC	anatomical therapeutic class
BOCF	baseline observation carried forward
BMI	body mass index
CC	Complete cases
C-CASA	Columbia- classification algorithm of suicide assessment
CI	confidence interval
CDR-SB	Clinical Dementia Rating – Sum of Boxes
CMH	Cochran-Mantel-Haenszel
CRF	case report form
CSD	Clinically significant difference
CSF	cerebrospinal fluid
CSR	clinical study report
C-SSRS	Columbia – Suicide Severity Rating Scale
CV	coefficient of variation
d_{Max}	Maximum effective dose
DMC	data monitoring committee
DSMB	data safety monitoring board
EAD	Early Alzheimer's Disease
ED_{90}	Dose regimen with at least 90% of the d_{Max} treatment effect
EPAD	European Prevention of Alzheimer's Dementia
FAQ	Functional Activities Questionnaire
FAS	full analysis set
IA	Interim Analysis
IMC	Interim Monitoring Committee
ITT	Intent to treat
LLOQ	Lower limit of quantitation
LOCF	last observation carried forward
LS	least squares

MAR	Missing at random
MCI	Mild Cognitive Impairment
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMSE	Mini Mental State Examination
MMRM	Mixed-effects models for repeated measures
NDLM	Normal Dynamic Linear Model
NI	non-inferiority
PET	positron emission tomography
PD	Pharmacodynamic
PK	Pharmacokinetic
PP	Per protocol
PT	Preferred term
RAR	Response adaptive randomization
SAE	serious adverse event
SAP	statistical analysis plan
SAS	statistical analysis system
SD	standard deviation
SE	standard error
SI	Système International
SOC	System organ class
TEAE	treatment-emergent adverse event
TEMAV	treatment-emergent markedly abnormal laboratory
TLG	tables, listings, and graphs
ULOQ	Upper limit of quantitation
vMRI	Volumetric magnetic resonance imaging
WHO	World Health Organization
WMS- LM	Wechsler Memory Scale – Logical Memory

3 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report results for the Core study of Eisai Protocol A Placebo-controlled, Double-blind, Parallel-group, Bayesian Adaptive Randomization Design and Dose Regimen-finding Study To Evaluate Safety, Tolerability and Efficacy of BAN2401 in Subjects With Early Alzheimer's Disease. The Extension Phase of this study will not be conducted as the conditions defined in study protocol were not met.

3.1 STUDY OBJECTIVES

3.1.1 Primary Objectives

1. To evaluate the efficacy of BAN2401 compared to placebo by establishing the ED_{90} (as defined in the protocol) for BAN2401 on the derived ADCOMS at 12 months of treatment in subjects with Early Alzheimer's Disease (EAD), defined as mild cognitive impairment (MCI) due to Alzheimer's disease (AD) – intermediate likelihood or mild Alzheimer's disease dementia

2. To assess the safety and tolerability of 3 doses and 2 dose regimens of BAN2401 in subjects with EAD

3.1.2 Key Secondary Objectives

1. To evaluate the effects of BAN2401 compared to placebo on brain amyloid pathophysiology at 18 months of treatment in subjects with EAD as measured by amyloid positron emission tomography (PET)

2. To evaluate the efficacy of BAN2401 compared to placebo on the ADCOMS at 18 months of treatment in subjects with EAD

3. To evaluate the efficacy of BAN2401 compared to placebo on the Clinical Dementia Rating – Sum of Boxes (CDR-SB) at 18 months of treatment in subjects with EAD

4. To evaluate the efficacy of BAN2401 compared to placebo on Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog) at 18 months in subjects with EAD
5. To evaluate the effects of BAN2401 compared to placebo at 18 months on clinical status separately within subjects with MCI and mild AD dementia for the following assessments: ADCOMS, CDR-SB, and ADAS-Cog
6. To evaluate the effects of BAN2401 compared to placebo on cerebrospinal fluid (CSF) biomarkers (Aβ[1-42], t-tau, and p-tau) at 18 months of treatment in subjects with EAD
7. To evaluate the effects of BAN2401 compared to placebo on total hippocampal volume using volumetric magnetic resonance imaging (vMRI) at 18 months of treatment in subjects with EAD

3.1.3 Secondary Objectives

1. To evaluate the effects of BAN2401 compared to placebo on brain amyloid pathophysiology at 12 months of treatment in subjects with EAD as measured by amyloid PET
2. To evaluate the effects of BAN2401 compared to placebo at 12 months on clinical status in subjects with EAD for the following assessments: ADCOMS, CDR-SB, and ADAS-Cog
3. To evaluate the effects of BAN2401 compared to placebo at 12 months on clinical status separately within subjects with MCI and mild AD dementia for the following assessments: ADCOMS, CDR-SB, and ADAS-Cog
4. To evaluate the effects of BAN2401 compared to placebo on cerebrospinal fluid (CSF) biomarkers (Aβ[1-42], t-tau, and p-tau) at 12 months of treatment in subjects with EAD
5. To evaluate the effects of BAN2401 compared to placebo on total hippocampal atrophy as measured by vMRI at 6 and 12 months, and left and right hippocampus, whole brain and total ventricular volume as measured by vMRI at 6, 12, and 18 months of treatment in subjects with EAD

3.1.4 Exploratory Objectives

1. To evaluate the effects of BAN2401 compared to placebo on clinical status across all clinical assessment time points not analyzed in Key Secondary and Secondary sections in subjects with EAD by assessment of:
    a. ADCOMS
    b. Clinical Dementia Rating – Sum of Boxes (CDR-SB)
    c. Mini Mental State Examination (MMSE)
    d. Functional Activities Questionnaire (FAQ)
    e. Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog)
2. To explore the relationship between change from baseline in clinical status at 12 and 18 months of treatment and baseline brain amyloid pathophysiology as measured by amyloid PET in subjects with EAD
3. To explore the relationship between change from baseline in clinical status and change from baseline in brain amyloid pathophysiology at 12 and 18 months of treatment in subjects with EAD as measured by amyloid PET
4. To explore efficacy of BAN2401 compared to placebo and the overall study population on the ADCOMS across all clinical assessment time points in Japanese subjects with EAD
5. To characterize the population pharmacokinetics (PK) of BAN2401 in EAD subjects, and to examine the effect of intrinsic and extrinsic factors on the PK

3.2 OVERALL STUDY DESIGN AND PLAN

This double-blind, parallel-group, placebo-controlled, multicenter and multinational study will utilize a dose-finding response adaptive randomization (RAR) design to evaluate the safety, tolerability, and efficacy of BAN2401 in subjects who have MCI due to Alzheimer's disease – intermediate likelihood, or who have mild Alzheimer's disease dementia. The study will take place over approximately 130 months at approximately 125 sites including North America and Europe, and Asia-Pacific countries. Approximately 3200 subjects will be screened to provide approximately 800 subjects to receive BAN2401 or matching placebo. In the event of early success in the Core Study at any interim analysis (IA) or at the 12 month Bayesian analysis, an OLE Phase will be implemented to allow for up to 60 months (5 years) of additional treatment.

The present study incorporates a Bayesian RAR design with frequent interim analyses to continually update randomization allocation on the basis of the primary clinical endpoint. This approach allows for ongoing assessment of drug futility or evidence for early success and for continued changes in randomization that favor efficacious treatment arms. Thus, the Bayesian approach not only limits exposure of subjects to non-efficacious treatment arms but can also mitigate the risks associated with larger and longer trials required to demonstrate clinical efficacy by leading to more efficient project termination or early advancement to a successful Phase 3 program. In the Core Study, after the 12-month assessments have been completed, treatment will continue to 18 months to follow the time course of any treatment effects observed at 12 months, and to evaluate biomarker and neuroimaging effects that may be consistent with potential disease modification.

Subjects will be from 2 clinical subgroups, collectively designated as EAD for the purposes of this protocol: (a) MCI due to AD – intermediate likelihood and (b) mild Alzheimer's disease dementia. At study entry, subjects will be stratified according to clinical subgroup, *APOE* status (*APOE4* positive and negative), and the presence or absence of ongoing AD treatment with AChEIs or memantine or both. Randomization into the 2 clinical subgroups (see Inclusion Criteria, Section 9.3.1 of the protocol) is to remain reasonably balanced whereby at least 60% of the total number of subjects will have MCI due to AD – intermediate likelihood clinical subgroup and at least 30% will have mild Alzheimer's disease dementia. The subjects will be randomized into 1 of 6 treatment groups (Table 1) as explained in Section 9.4.4 and Section 9.4.5 of the protocol. *APOE4* status (positive or negative) must be confirmed for all subjects prior to randomization. After Protocol Amendment 05 based on the request from the Voluntary Harmonisation Procedure (VHP) committee, subjects who are confirmed *APOE4* positive (*APOE4* hetero- or homozygous) will not be randomized to the 10 mg/kg, biweekly dose. Randomization in the Core Study will initially be fixed to each of these dose groups (4:2:2:2:2:2 ratio, placebo to each of the 5 active treatment arms). All study personnel and subjects will be blinded with respect to the dose regimens. After 196 subjects have been randomized into the study, an interim analysis (IA) will be conducted and RAR will guide subsequent randomization into dose groups. Interim analyses and RAR will be repeated after 250 subjects have been randomized and again after each additional 50 subjects until all 800 subjects are randomized, and will then be conducted at 3-month intervals until all subjects complete 12 months of

treatment. A Bayesian analysis will be conducted at 12 months of treatment to assess for early success, and at 18 months of treatment, regardless of whether early success is achieved at 12 months.

Table 1 BAN2401 Dose Regimens

Study Arm	BAN2401 Dose (mg/kg)	Infusion frequency
1	Placebo	Biweekly[a]
2	2.5	Biweekly[a]
3	5.0	Biweekly[a]
4	10.0[b]	Biweekly[a]
5	5.0	4-week intervals[c]
6	10.0	4-week intervals[c]

a: Biweekly = 2-week intervals
b: Subjects who are confirmed *APOE4* positive (*APOE4* hetero-or homozygous) will not be randomized to the 10 mg/kg, biweekly dose.
c: Subjects receiving study drug at 4-week intervals will receive placebo at intervening 2-week time points.

An overview of the study design is presented in Figure 1.

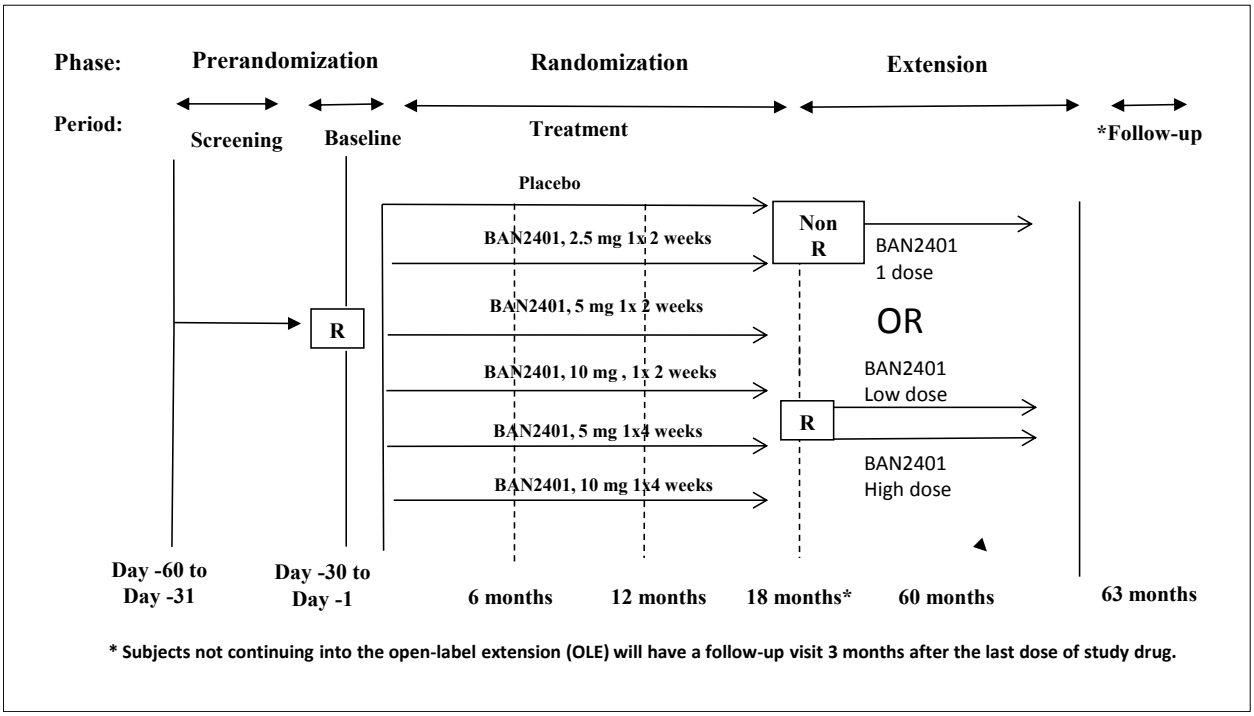

Figure 1 Design of Study BAN2401-G000-201
R = randomization.

Oversight of the interim analyses and RAR will be conducted by an independent Interim Monitoring Committee (IMC). The IMC will consist of 3 members who are external to the company with expertise in Bayesian adaptive design in clinical trials.

The interim analyses and RAR will be conducted by an unblinded external independent data analysis group in accordance with the protocol. This independent group will provide the interim analysis outcomes to the IMC at each interim look. The IMC will ensure the integrity of the IA and response adaptive randomization process through review of primary efficacy data. In addition, the IMC will act in an independent advisory capacity to monitor the Bayesian IA outcomes according to the early success and futility boundaries pre-specified in the protocol. The IMC will inform the sponsor if the boundaries for early success or futility have been crossed. If the study continues to randomize subjects after statistical futility boundary confirmed by the IMC is crossed, it will be considered a failed study. The IMC will not be charged with any subject safety issues, as this will be the responsibility of an independent DSMB.

A separate unblinded DSMB will be employed at regular intervals to monitor the overall safety of the Core Study and OLE Phase of the study and will make recommendations to the Sponsor as appropriate. In addition, the DSMB will meet when a predetermined number of subjects are randomized to assess whether the option for home infusions is feasible (per DSMB charter, if allowed and conducted according to country and local guidelines; home infusions will not be allowed in Germany). Upon implementation of Protocol Amendment 07, newly enrolled subjects will not be offered the option for home infusions. Subjects opting for home infusions before implementation of Protocol Amendment 07 will be allowed to continue with home infusions for the duration of their participation in the study. Details regarding DSMB membership and operational characteristics will be provided in the DSMB Charter.

If the statistical futility boundary as confirmed by the IMC is crossed, the IMC will inform the DSMB. The DSMB will make risk-benefit assessments after reviewing emerging safety and efficacy signals and will provide its recommendations to the sponsor. The sponsor will make the final decision pertaining to study futility after reviewing recommendations by the DSMB.

Administration of BAN2401 to an individual subject may be stopped on the basis of safety-related findings as outlined in Section 9.3.3 of the protocol.

The scheduled of assessments are presented in Tables 6 and 7 of the protocol.

4 DETERMINATION OF SAMPLE SIZE

The sample size and design characteristics to test the hypothesis under the primary objective based on the primary endpoint, the ADCOMS at 12 months, were determined by means of simulations. Extensive trial simulations have shown that a total of 800 subjects will be sufficient to demonstrate that the most likely ED_{90} dose achieves a clinically significant difference (CSD) from placebo with a probability of at least 95% in the interim analyses and greater than 80% in the analysis at 12 months if the trial does not stop for early

success. With 800 subjects, the average probability of study success is 80% across a wide range of dose response scenarios, and there is at least 80% probability of study success for the dose response scenarios where the treatment is considered to show a clinically significant difference from placebo as defined in the Protocol. Simulations have also shown that if there is no efficacy at all for any dose, then the probability of falsely claiming superiority to placebo is no more than 10% assuming a 20% dropout rate.

For each of the 6 treatment groups (5 active dose regimens and placebo) the final number of subjects per group will differ depending on the observed interim treatment responses. The simulation plan is described in the appendix section of this document which presents further details.

vMRI

The null hypothesis is that there is no difference between active dose and placebo. There are 5 null hypotheses corresponding to 5 active dose regimens. The alternative hypothesis is that at least 1 null hypothesis is false (1-sided). The null hypotheses will be tested using Dunnett-Hsu method with 1-sided alpha of 0.05. The statistical power was estimated through simulation for a moderate dose-response assumption that the percent reduction in change from baseline of total hippocampal volume compared to placebo would be 15%, 20%, 25%, 15%, and 20% corresponding to the 5 dose regimens (2.5, 5, and 10 mg/kg biweekly, 5 and 10 mg/kg 4-week interval), respectively. The estimated power for actual planned study sample size is 76.4% at 12 months and 69.4% at 18 months, assuming an attrition rate of 20% at 12 months and an exponential dropout model. Under a stronger dose-response assumption (ie, the percent reduction in change from baseline of total hippocampal volume compared to placebo is twice as much as that in the moderate dose-response assumption), the estimated power is at least 99% at both 12 and 18 months.

AMYLOID PET (IMAGING SUBGROUP)

The null and alternative hypotheses and statistical test are the same as that for vMRI. To estimate the sample size for amyloid PET imaging subgroup at 18 months, a common standard deviation of change from baseline was estimated as 0.4 and a reasonable mean difference between treatment and placebo was estimated as 0.25 (SUVR). The corresponding standard deviation and mean difference at 12 months were estimated as 0.27 and 0.17, respectively. Assuming a moderate dose-response that the best dose regimen would achieve a difference of 0.25 (SUVR), and that the 2 middle and 2 low dose regimens would achieve a difference of 0.2 and 0.15 (SUVR), respectively, it would require a total of 230 subjects from all 6 arms at 18 months to achieve an 80% power. This sample size took into account possible unequal number of subjects per arm. Under the same attrition rate assumption as that for vMRI, a total sample size of 306 is required at 12 months in order to target 230 subjects at 18 months. The estimated power with 306 subjects at 12 months is approximately 85%. Because amyloid PET imaging is required for every subject at baseline, the post baseline amyloid PET in the imaging subgroup will be planned for the first 306 subjects who have consented and who are still on treatment at 12 months.

5 STATISTICAL METHODS

In this Phase 2 dose-finding study of BAN2401 for the treatment of Early Alzheimer's Disease, response adaptive randomization will be used to allocate subjects to placebo control or 1 of the 5 active doses with the goal of characterizing the dose response. The study will be monitored for early success and early futility. The sponsor intends that the futility rule for stopping will be binding. If the study continues to randomize subjects after futility boundary has been crossed, it will be considered a failed study.

All statistical analyses for Core Study will be performed by the sponsor or designee. All of the Bayesian analysis will be conducted by an external IMC. The sponsor and investigators will be blinded to the IA data and outputs. The detailed processes and procedures related to study success at IA and the Bayesian analysis at 12 months are defined in a separate integrity charter. All other statistical analyses for Core Study will be performed by the sponsor or designee after the database is locked and released for unblinding. At each IA of the primary endpoint, a snapshot of the data will be obtained and released for analysis. A copy of this snapshot will be archived. This process will also be followed for the safety data that will be used for analysis for the DSMB. Statistical analyses will be performed using SAS software or other validated statistical software as required. All statistical tests for conventional analysis will be based on the two-sided 10% or one-sided 5% level of significance. The definition of success at primary analysis based on Bayesian methods is given in Section 5.4.2 Adaptive Trial Definitions.

All descriptive statistics for continuous variables will be reported using mean, standard deviation (SD), median, minimum and maximum. Categorical variables will be summarized as number (percentage) of subjects.

5.1 STUDY ENDPOINTS

5.1.1 Primary Endpoint(s)

- Change from baseline in the derived ADCOMS at 12 months

5.1.2 Key Secondary Endpoint(s)

- Change from baseline at 18 months in brain amyloid pathophysiology as measured by amyloid PET
- Change from baseline in the derived ADCOMS at 18 months
- Change from baseline in CDR-SB at 18 months
- Change from baseline in ADAS-Cog at 18 months
- Change from baseline in CSF biomarkers (Aβ[1-42], t-tau, and p-tau) at 18 months
- Change from baseline in total hippocampal volume at 18 months using vMRI

5.1.3 Secondary Endpoint(s)

- Change from baseline at 12 months in brain amyloid pathophysiology as measured by amyloid PET
- Change from baseline at 12 months on clinical status for the following assessments: ADCOMS, CDR-SB, and ADAS-Cog

- Change from baseline in CSF biomarkers (Aβ[1-42], t-tau, and p-tau) at 12 months
- Change from baseline in total hippocampal volume at 6 and 12 months, and in left and right hippocampus, whole brain, and total ventricular volume as measured by vMRI at 6, 12, and 18 months

5.1.4 Exploratory Endpoint(s)

- Change from baseline in clinical status across all time points not analyzed in Key Secondary and Secondary sections for ADCOMS, CDR-SB, ADAS-Cog, MMSE, and FAQ

5.2 STUDY SUBJECTS

5.2.1 Definitions of Analysis Sets

The number (percent) of subjects included in each analysis set will be presented by analysis set, treatment group and overall. The description of the analysis data sets are:

The Randomized Set is the group of subjects who are randomized to study drug.

The Safety Analysis Set is the group of subjects who receive at least 1 dose of study drug and have at least 1 post dose safety assessment.

The Full Analysis Set is the group of randomized subjects who receive at least 1 dose of study drug and have baseline and at least 1 post dose primary efficacy measurement.

The PK Analysis Set is the group of subjects with at least 1 quantifiable BAN2401 serum concentration with a documented dosing history.

The PD Analysis Set is the group of subjects who have sufficient PD data to derive at least 1 PD parameter.

The Per Protocol (PP) Analysis Set is the subset of subjects in the Full Analysis Set who complied with the protocol.

5.2.2 Subject Disposition

The number of subjects screened, the number (percent) of subjects who failed screening, and the reasons for screen failure will be summarized, based on data reported on the Screening Disposition eCRF. The distribution of the number of randomized subjects by each site will be summarized for each randomized treatment group. The primary reasons for screen failures; did not meet inclusion or met exclusion criteria, AE, lost to follow-up, withdrawal consent, and other, will be presented.

Study Completion: the number (percent) of randomized and treated subjects who completed the study and who discontinued from the study will be summarized according to the primary

reason for discontinuation and secondary reason(s) for discontinuation, based on data reported on the Subject Disposition (Study Phase) CRF. The number (percent) will be presented by treatment group and total for subjects; randomized, not treated, treated, who completed the study, and discontinued from the study. The primary reasons for discontinuation from the study are; AE, lost to follow-up, subject choice, withdrawal of consent, pregnancy, study terminated by sponsor, and other. The secondary reasons for discontinuation from the study are: AE, subject choice, and other.

Completion of Study Treatment: the number (percent) of randomized and treated subjects who completed study treatment and who discontinued from study treatment will be summarized according to the primary reason for discontinuation and also according to secondary reason(s) for discontinuation, based on data reported on Early Discontinuation from Study Drug eCRF. The number (percent) will be presented by treatment group and total for subjects; randomized, not treated, treated, who completed study treatment, and discontinued from study treatment. The primary reasons for discontinuation from the study treatment are: AE, subject choice, inadequate therapeutic response, pregnancy, and other. The secondary reasons for discontinuation from the study are: AE, subject choice, inadequate therapeutic effect, and other. Subjects who discontinued study treatment but were followed up for efficacy assessments post treatment will also be summarized.

5.2.3 Protocol Deviations

A listing of subjects with protocol deviations will be provided by subject along with the description of the protocol deviation. Protocol deviations will be identified prior to database lock.

5.2.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics for the Safety Analysis set and Full Analysis set will be summarized for each treatment group and baseline PET SUVR subgroup defined in Section 5.8 using descriptive statistics. Continuous demographic and baseline variables include age, ADCOMS, ADAS-Cog, CDR-SB, FAQ, and MMSE; categorical variables include age < 65, age ≥ 65 to <80, age ≥ 80, sex, race, ethnicity, region, Global CDR score, memory box score of the CDR, *APOE ε4 carrier status* (positive (overall, heterozygous and homozygous) and negative), ongoing treatment with AChEIs and/or memantine (yes or no), and clinical subgroup (MCI due to AD and Mild Alzheimer's Disease Dementia).

The status of the disease at screening for each clinical subgroup (MCI due to AD and Mild Alzheimer's Disease Dementia) will be summarized according to the number of years the subject had the disease since diagnosis and onset of symptoms, his/her age at diagnosis, and age at onset of symptoms (summarized both as a continuous variable and as a categorical variable, with categories 30 to < 50 years, ≥ 50 to < 65 years, ≥ 65 years to < 80, and ≥ 80).

MEDICAL HISTORY

The medical history verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). Medical history will be coded to the MedDRA (Version 15.0 or higher) lower level term (LLT) closest to the verbatim term. The linked MedDRA preferred term (PT) and primary system organ class (SOC) will also be captured in the database.

The number (percent) of subjects will be presented by SOC and PT for each treatment group.

5.2.5 Prior and Concomitant Therapy

All investigator terms for medications recorded in the CRF will be coded to an 11-digit code using the World Health Organization Drug Dictionary (WHO DD). The number (percentage) of subjects who took prior and concomitant medications will be summarized on the Safety Analysis Set by treatment group, Anatomical Therapeutic Chemical (ATC), and WHO DD preferred term. Prior medications are defined as medications that stopped before the first dose of study drug. Concomitant medications are defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug and continued up to the last dose. Prior and concomitant therapies for AD and those for non-AD will be summarized separately.

5.2.6 Treatment Compliance

Treatment compliance rate is defined as 100% x (total number of active infusions subjects actually received)/(total possible number of active infusions the subjects could receive). For example, for a subject assigned to the bi-weekly arms, the total possible number of active infusions the subject could receive will be 26 or 39 over 12 or 18 months of treatment period, respectively; for a subject assigned to the monthly arms, the total possible number of active infusions the subject could receive will be 12 or 18 over 12 or 18 months of treatment period, respectively. Treatment compliance rate will be summarized by treatment group both as a continuous variable and as a categorical variable, with categories <50%, 50% to 75%, 75% to 100%. Subjects in bi-weekly arms with <50% compliance rate and subjects in monthly arms with <75% compliance rate will be excluded from Per Protocol Set.

5.3 DATA ANALYSIS GENERAL CONSIDERATIONS

5.3.1 Pooling of Centers

This study is a multicenter, international study with approximately 125 sites. Sites will be pooled into the following regions, in order to have sufficient number of subjects per treatment group:

- North America (USA & Canada)
- Western Europe (including the Oceana region)
- Asia (including Japan)

5.3.2 Adjustments for Covariates of Interest

The primary, secondary, and exploratory endpoints will include the stratification variables clinical subgroup (MCI due to AD, Mild Alzheimer's Disease Dementia), *APOE ε4 carrier status* (positive or negative) and the presence or absence of ongoing AD treatment (i.e., AChEIs and/or memantine). In addition, the following covariates of interest will be evaluated in the statistical models: Age (treated as continuous and as categorical age ≤ 65, 65 < age < 80, age ≥ 80 years), region and *APOE* alleles (ε4/ε4, ε4/other and other/other)

5.3.3 Multiple Comparisons/Multiplicity

This study has one primary endpoint so no adjustment for multiplicity is required. The primary endpoint will be analyzed using a Bayesian Normal Dynamic Linear Model; therefore, no adjustment for multiplicity is required. Key secondary analyses compare the combined 10mg treatment groups with placebo, so there is no need to adjust multiplicity for multiple dose groups.

5.3.4 Examination of Subgroups

The primary endpoint, key secondary endpoints, secondary endpoints and exploratory endpoints related to clinical assessments and biomarkers will be analyzed using the mixed-effects model with repeated measures to evaluate the following subgroups; age ≤ 65, 65 < age < 80, age ≥ 80, gender, ethnicity, race, region, clinical subgroup (MCI due to AD, Mild Alzheimer's Disease Dementia), *APOE ε4* carrier status (positive (overall, heterozygous and homozygous) or negative), and the presence or absence of ongoing AD treatment (i.e., AChEIs and/or memantine), subjects with/without neutralizing ADA, and biomarker subgroups defined in Section 5.8.

5.3.5 Handling of Missing Data, Drop-outs, and Outliers

The missing primary endpoint will be imputed using Bayesian imputation method based on longitudinal model in the primary Bayesian analysis (Section 5.4.2.2). In the conventional analyses, the missing primary endpoint will be handled by the Mixed-effects Models for Repeated Measures (MMRM). Other statistical methods for missing data in conventional analyses may also be performed as sensitivity analyses. Missing values in key secondary endpoints, secondary endpoints and exploratory endpoints related to clinical assessments and biomarkers will be handled by the MMRM in the conventional analyses.

Efficacy Variables

If any item is missing within the Composite Score (ADCOMS), ADAS-cog, MMSE, CDR, and FAQ then their respective total scores will be missing.

CSF Biomarkers

Subjects with baseline values ≥ ULOQ or ≤ LLOQ will be excluded from all statistical analyses.

For analyses defined in Section 5.8, only subjects with post-baseline values > LLOQ and < ULOQ will be included. For additional analyses, subjects with at least 1 post-baseline value,

including those ≥ ULOQ or ≤ LLOQ, will be categorized as follows: "stable", "increased from baseline", and "decreased from baseline". For subjects that have numerical differences calculated, any increases will be categorized as "increased from baseline" and any decreases will be categorized as "decreased from baseline" regardless of magnitude. Difference of 0 will be categorized as "stable". Subjects with baseline values > LLOQ and < ULOQ and post-baseline values ≥ ULOQ will be categorized as "increased from baseline", subjects with baseline values > LLOQ and < ULOQ and post-baseline values ≤ LLOQ will be categorized as "decreased from baseline".

When an improved assay becomes available, all samples may be re-run using the improved assay.

Adverse Events

Adverse events with missing severity will be assigned the highest severity. Adverse events with missing relationship will be assigned the highest relationship.

5.3.6 Other Considerations

Not Applicable.

5.4 EFFICACY ANALYSES

5.4.1 Primary Efficacy Analyses

The primary analysis will be based on subjects from the Full Analysis Set with the prespecified censoring rules applied. The primary endpoint is the change from baseline to 12 months in the derived ADCOMS. The derived ADCOMS is a weighted linear combination of 12 items from the three existing clinical scales, the ADAS-cog, the MMSE, and the CDR. These 12 items consist of the predictive variables A4, A7, A8, A11, M1, M7, C1, C2, C3, C4, C5 and C6, which have been selected from the clinical scales, the ADAS-cog, the MMSE, and the CDR. The names of these item and the corresponding scale names are described in Table 2.

Table 2: Predictive Variables for the Derived ADCOMS

Scale	Item ID	Item Name	PLS weight
ADAS-cog	A4	Delayed Word Recall	0.00847483
	A7	Orientation	0.017088
	A8	Word Recognition	0.003732761
	A11	Word Finding	0.016211
MMSE	M1	Orientation Time	0.041567
	M7	Drawing	0.038238
CDR	C1	Personal Care	0.054321
	C2	Community Affairs	0.1091
	C3	Home and Hobbies	0.089039
	C4	Judgment and Problem Solving	0.069493
	C5	Memory	0.058724
	C6	Orientation	0.078152

ADAS-cog = Alzheimer's Disease Assessment Scale, cognitive subscale, CDR = Clinical Dementia Rating, ID = identification, MMSE = Mini Mental State Examination, PLS = Partial Least Squares.

The derived composite score will be calculated from these predictive variables according to the formula:

Derived composite Score =A4 *0.00847483 + A7 *0.017088 + A8 *0.003732761 + A11*0.016211 + (5-M1)*0.041567 + (1-M7)*0.038238 + C1*0.054321+ C2*0.1091+ C3*0.089039 + C4*0.069493 + C5*0.058724 + C6*0.078152.

The maximum derived composite score is achieved when each item is assigned the maximum score. This maximum composite score is 1.97. The range of this new composite score is therefore between 0 and 1.97.

The primary endpoint is the change from baseline to 12 months in the derived ADCOMS. The primary efficacy analysis is based on Bayesian statistics. The dose-response of the primary endpoint is modeled with a two-dimensional first-order normal dynamic linear model (NDLM), where Normal and Inverse-Gamma priors are used. The primary efficacy analysis will calculate the Bayesian posterior probability that the dose identified is the most likely ED_{90} dose that achieves the CSD compared to the placebo arm. At each analysis, three Bayesian probabilities will be summarized for each active dose: the probability of being the ED_{90} dose, the probability of being statistically superior to placebo, and the probability of being statistically superior to placebo by the CSD. The study will be considered a success in the Bayesian analysis at the completion of 12 months of treatment if there is at least an 80% probability that the ED_{90} achieves the clinically significant difference from placebo. See Section 5.4.2 for definitions of CSD, ED_{90}, early success and futility.

Censoring

For the clinical efficacy data (i.e. derived ADCOMS, MMSE, CDR-SB, ADAS-cog and FAQ)

- Subjects will be censored at the time of initiation of new AChEIs or memantine treatment regimens if they were not on AChEIs or memantine at randomization, and will be censored at the time of dose adjustment of AChEIs or memantine if they were already on stable treatment with AChEIs or memantine at randomization.

The value of the primary endpoint for censored subjects and censored assessments will be imputed using data up to the censoring time and the Bayesian imputation method.

5.4.1.1 Sensitivity Analyses

After unblinding, the primary endpoint will be analyzed using the Bayesian methods described above. The following additional Bayesian analyses will be conducted using the Full Analysis Set:

- The primary endpoint will be analyzed regardless of initiation of new AChEIs or memantine or dose adjustment of stable AChEIs or memantine, using the same Bayesian analysis method as in the primary analysis. This is an intent-to-treat analysis without any censoring.

- In case of early success, the primary analysis as well as the above sensitivity analysis will be repeated after all subjects have either completed 12 months follow-up or are lost to follow-up.

5.4.1.2 Conventional Analyses

Statistical methods for the conventional analyses of primary efficacy endpoint will use a mixed effects model with repeated measures (MMRM) comparing placebo to the identified ED_{90} dose from the Bayesian analysis. The model will include baseline ADCOMS as covariate, with treatment group, visit, randomization stratification variables (i.e., clinical subgroup [MCI due to AD, Mild Alzheimer's Disease Dementia], the presence or absence of ongoing AD treatment [AChEIs or memantine or both], *APOE4* status [positive, negative]), region, treatment group-by-visit interaction as fixed effects. If there is an imbalance across treatment groups based on randomization stratification variables, additional MMRM model will be run without the randomization stratification variables. In addition, ANCOVA model will also be used with treatment as a factor and baseline value as a covariate.

An unstructured covariance matrix will be employed in the MMRM to model the covariance of within subject effect. These analyses will be performed using all randomized subjects in the full analysis set as well as in the per protocol set. There will be no conventional analysis at each interim analysis before early success is claimed and as such there will be no multiplicity adjustment made to account for the interim analyses. In case of early success, the conventional analyses will be performed using the data collected prior to the interim analysis as well as at the end of 12 months assessment, and the analysis at the end of 12 months assessment will be considered as sensitivity analysis to the analysis at early success.

5.4.2 Adaptive Trial Definitions

Clinically Significant Difference (CSD): The CSD used in the adaptive model relative to the change from baseline over time for placebo in the derived ADCOMS. The CSD for this study represents an approximate 25% reduction compared to placebo in the progression of AD as measured by the derived ADCOMS.

Maximum Effective Dose (d_{Max}): The dose that achieves the greatest treatment effect.

ED_{90}: The 'simplest' dose regimen that achieves at least 90% of the d_{Max} treatment effect. The term 'simplest' reflects the preference for dosing frequency and amount when considering the treatment arms for the study and while taking efficacy into account (monthly administration is preferred over bi-weekly administration and lower doses are preferred over higher doses, with priority placed on frequency of administration). Thus, the rank order of prioritized doses (beginning with the simplest) is: 5 mg/kg Q4wk, 10 mg/kg Q4wk, 2.5 mg/kg Q2wk, 5 mg/kg Q2wk, and 10 mg/kg Q2wk.

Early Success: The study meets early success criteria if the probability that the ED_{90} at 12 months of treatment is better than the placebo by the CSD is at least 0.95. Early success is achieved before all subjects complete 12 months of treatment and can only be determined once ≥350 subjects are randomized. If early success is declared prior to full randomization, the randomization of new subjects will stop. Subjects already randomized into the study will receive 18 months of treatment.

Futility: The study meets statistical futility criteria if the probability that the ED_{90} is better than the placebo by the CSD is less than 0.05 (with ≤300 subjects randomized) or 0.075 (with ≥350 subjects randomized) at 12 months of treatment. If the study meets statistical futility criteria, the sponsor will make the final decision pertaining to study futility after reviewing recommendations by the DSMB. Based on Sponsor's final decision, if the study continues to randomize subjects, it will be considered a failed study.

Study Success: The study will be considered a success if either of the following criteria are met for the 12 month treatment endpoint:

1. The study meets early success criteria (as outlined above)

2. The study reaches complete randomization and the Bayesian analysis at completion of 12 months of treatment results in an ED_{90} with at least an 80% probability of being better than the placebo by the CSD

Further details are presented in the Simulation Plan, see Appendix 13.2.

5.4.2.1 Dose-Response Model

The dose-response of the primary endpoint is modeled with a two-dimensional first-order normal dynamic linear model (NDLM), where Normal and inverse-Gamma priors are used. The primary efficacy analysis will calculate the Bayesian posterior probability (the conditional probability

when taking the observed data and prior efficacy information into account) that the dose identified as the most likely ED_{90} achieves the CSD compared to the placebo arm. The probability is obtained using the Markov Chain Monte Carlo technique and the posterior distributions of the parameters in the dose response model. The posterior distribution is the distribution of parameter conditional on the data observed from the study. A detailed description of the NDLM and simulations is provided in the Simulation Plan, Appendix 13.2 of this document.

5.4.2.2 Longitudinal Model

At each interim analysis there will be subjects who have complete or incomplete information. Some subjects will have complete information up to and including 12 months. Other subjects will have earlier observations, but no 12-month observations. There will be subjects with no observations.

The information from subjects with incomplete observations is utilized to the extent that the earlier values are correlated to the final 12-month value. A Bayesian model is built to learn from the accruing information on the association between the early values, and the final endpoint at 12 months. Further details on the longitudinal model are presented in the Simulation Plan, see Appendix 13.2.

5.4.3 Report of Bayesian Analysis of Primary Endpoint in Adaptive Trial

The Bayesian analysis of primary endpoint as described in Section 5.4.1 will be performed by an independent data analysis group (Tessella) and the analysis results will be reported in a separate document, IMC (Interim Monitoring Committee) report, at time of success at IA and the Bayesian analysis at 12 months. The following analysis outputs will be provided in this report:

1. Area chart plot of number of subjects randomized over time by treatment group.
2. Scatter plot of primary endpoint along with NDLM estimate of mean ± 2 SD by treatment group. Each subject will be plotted with marker type and color identifying the last visit in case some subjects do not have 12 months visit.
3. Summary of posterior estimate of mean and standard deviation of primary endpoint for each treatment group.
4. Plot of NDLM estimates of primary endpoint (mean ± 2 SD) showing Bayesian dose-response relationship based on the two-dimensional first-order NDLM.
5. Summary of posterior probabilities of being dMax and ED_{90}, superiority to placebo, and superiority to placebo by CSD for each dose.
6. Column chart plot of Bayesian posterior probabilities as number 5 above for each dose.

5.4.4 Secondary Efficacy Analyses

The Full Analysis Set will be used in secondary efficacy analyses.

Key Secondary Endpoint Analyses

Key Secondary endpoints and corresponding analyses are ordered to reflect the most recent FDA draft guidance (February, 2018) and EMA Final Guideline (February, 2018) that prioritize effects on pathophysiology in early AD, as measured by biomarkers. Conventional analysis will be performed using all available data at the time of success (if applicable), and at 18 months of treatment, regardless of success. As described in section 5.4.1.2, conventional analysis will be based on the MMRM model including baseline as covariate, with treatment group, visit, randomization stratification variables (i.e., clinical subgroup [MCI due to AD, Mild Alzheimer's Disease Dementia], the presence or absence of ongoing AD treatment [AChEIs or memantine or both], *APOE4* status [positive, negative]), region, treatment group-by-visit interaction as fixed effects. If there is an imbalance across treatment groups based on randomization stratification variables, additional MMRM model will be run without the randomization stratification variables. In addition, ANCOVA model will also be used with treatment as a factor and baseline value as a covariate. These statistical models will be used to compare the combined 10 mg dose group (including bi-weekly and monthly regimens) with placebo for the key secondary endpoints. The rationale for combining the two 10 mg dose regimens is to account for the loss of subjects positive for APOE4 in the 10 mg/kg biweekly dose group and the inability to randomize APOE4 carriers to the 10 mg/kg biweekly group following a regulatory request by European Health Authorities in July 2014.

In addition, analysis will be performed for the following treatment comparisons:

- Combining 2 high doses (10 mg bi-weekly + 10 mg monthly), 2 middle doses (5 mg bi-weekly + 5 mg monthly), and the 2.5 mg bi-weekly dose, resulting in 3 treatment comparisons with placebo (for dose response)
- ED_{90} dose of BAN2401 versus placebo, where the ED_{90} dose of BAN2401 will be established by Bayesian analysis at 18 months as described in the next paragraph
- By treatment regimen (10 mg bi-weekly, 10 mg monthly, 5 mg bi-weekly, 5 mg monthly, 2.5 mg bi-weekly, placebo)
- Combining the top 3 doses (10 mg bi-weekly, 10 mg monthly, and 5 mg bi-weekly) versus placebo

The key Secondary endpoints (ADCOMS, CDR-SB, and ADAS-Cog) will also be analyzed separately within subjects with MCI and mild AD dementia.

Change from baseline in the ADCOMS at 18 months will be analyzed, as a sensitivity analysis to the conventional analysis, using the same Bayesian methodology as that for analysis of change from baseline in the ADCOMS at 12 months, but using the full 18 months of efficacy data with the model projecting to 18 months of treatment. The Bayesian analysis of change from baseline in the ADCOMS at 18 months is positive if the analysis results in an ED_{90} dose with at least an 80% probability of being better than the placebo by the CSD. The ED_{90} dose will be identified based on change from baseline in the ADCOMS at the 18 month Bayesian analysis and will be used to compare ED_{90} dose of BAN2401 and placebo in the above sensitivity analyses for all key secondary endpoint analyses.

Specifically, the following Bayesian analyses will be performed at 18 months:

- Fit the 2-dimensional NDLM Bayesian model defined in Appendix 13.2 with 5 doses across 2 schedules of administration for the 18-month endpoint. This includes imputation of missing values from the longitudinal model.
- Compare placebo versus the combined 10mg dose group. This includes imputation of missing values from the longitudinal model.
- Fit a one-dimensional Bayesian model with 3 doses: 2.5mg dose group, combined 10mg dose group, and combined 5mg dose group. This will be a one-dimensional NDLM model with imputation of missing values from the longitudinal model.
- Fit the European Prevention of Alzheimer's Dementia (EPAD) disease progression model for selected doses versus the placebo control.

In this study, subjects consented to PET sub-study have been using two different tracers for PET scan: Florbetapir and Flutemetamol. Only less than 10 subjects have used Flutemetamol and the rest have used Florbetapir. It is possible that different tracers may yield different results and data using different tracers may not be directly comparable, therefore only subjects using Florbetapir will be included in summary tables and other statistical analysis based on PET data. Subjects using Flutemetamol will only be listed and will not be included in summary tables or other statistical analysis based on PET data.

Other Secondary Endpoint Analyses

The same analysis described above for key secondary endpoints will be performed for other secondary endpoints.

For vMRI and amyloid PET endpoints, the adjusted p-value based on Dunnett-Hsu method with 1-sided alpha of 0.05 will also be provided in addition to the p-value corresponding to pairwise comparison.

5.4.5 Other Efficacy Analyses

No other efficacy analyses are planned for this study.

5.5 Pharmacokinetic, Pharmacodynamic, and PHARMACOGENOMIC/PHARMACOGENETIC ANALYSES

Further details for the analyses described below will be presented in a separate analysis plan.

5.5.1 Pharmacokinetic Analyses

The Safety Analysis Set will be used for individual BAN2401 serum and CSF concentration listings. The PK Analysis Set will be used for summaries of BAN2401 serum and CSF concentrations.

A population PK approach will be used to characterize the PK of BAN2401. The effect of covariates (i.e., demographics) on BAN2401 PK will be evaluated. The PK model will be parameterized for clearance (CL) and volumes of distribution. Derived exposure parameters such

as AUC will be calculated from the model using the individual posterior estimate of CL and dosing history.

5.5.2 Pharmacodynamic Analyses

The PD Analysis Set will be used for the summaries and analyses of CSF biomarkers. CSF Aβ(1-42), t-tau, and p-tau will be assessed and data presented graphically. Baseline levels and changes in t-tau and p-tau will be correlated with changes in CSF Aβ(1-42), imaging markers, and *APOE4* status. Summary statistics for CSF biomarkers will be assessed for evidence of a dose relationship.

5.5.3 Pharmacokinetic/Pharmacodynamic

The PK/PD relationship between CSF biomarker levels and serum PK parameters or CSF concentrations of BAN2401 will be explored graphically, and any emergent relationship will be explored through population PK/PD modeling. These serum PK parameters include SS, C_{max} and $AUC_{(0\text{-}t)}$ derived from the population PK model. The PK/PD relationship between serum PK parameters and CSF concentrations of BAN2401 with other biomarkers may also be explored using similar methods.

Additionally, the relationship between various serum PK parameters (e.g. C_{max}) or CSF concentrations of BAN2401 and the derived ADCOMS at 12 months, and the relationship between various PK exposure parameters or CSF concentrations of BAN2401 and the change from Baseline for 12 and 18 months in Modified ADAS-cog, CDR-SB and MMSE, will be explored graphically. Any emergent relationships will be explored through population PK/PD modeling.

The relationship between exposure to BAN2401 and most frequent AEs will also be explored.

5.5.4 Pharmacogenomic/Pharmacogenetic Analyses

APOE4 genotypes will be used in the statistical analysis to determine the effects on treatment response and safety.

Exploratory pharmacogenomics and biomarker analyses may be performed and reported separately. Details of these analyses may be described in a separate analysis plan.

5.6 SAFETY ANALYSES

Evaluations of safety will be performed on the Safety Analysis Set. The incidence of AEs (including changes from baseline in physical examination), out-of-normal-range laboratory safety test variables, abnormal ECG findings, anti-BAN2401 antibody titers, and out-of-range vital signs, suicidality (C-SSRS), along with change from baseline in laboratory safety test variables, ECGs, safety MRI, and vital sign measurements, will be summarized by treatment group using descriptive statistics. The safety analysis will be also be summarized by the following safety subgroups: age (≤65, 65 < age < 80, ≥ 80 years), sex (male, female), race (white, black, Asian, other), region, *APOE4* status, and clinical subgroup (MCI due to AD – intermediate likelihood or mild Alzheimer's disease dementia).

Study Day 1 for all safety analyses is defined as the date of the first dose of study drug.

5.6.1 Extent of Exposure

Extent of exposure to study drug will be summarized as cumulative extent of exposure from Day 1 to Week 6 initially, and at 3 months increments thereafter. The cumulative number and percent of subjects will be presented by treatment group. Duration of exposure will be summarized using descriptive statistics and by duration categories; Day 1 to 6 weeks, 6 weeks to 3 months, and 3-month intervals thereafter. The number and percent of subjects will be presented by each duration category and treatment group. Overall exposure (number of subject-months) is defined as summation over all subjects' exposure durations and will be summarized by treatment group.

5.6.2 Adverse Events

The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to the MedDRA (Version 15.0 or higher) lower level term (LLT) closest to the verbatim term. The linked MedDRA preferred term (PT) and primary system organ class (SOC) will also be captured in the database.

A treatment-emergent AE (TEAE) is defined as an AE that emerged during treatment or within 90 days following the last dose of study drug, having been absent at pretreatment (Baseline) or

- Reemerged during treatment, having been present at pretreatment (Baseline) but stopped before treatment, or
- Worsened in severity during treatment relative to the pretreatment state, when the AE was continuous.

TEAEs and AEs will be followed for the entire duration of the study. Only those AEs that are treatment emergent will be included in summary tables. All AEs, treatment emergent or otherwise, will be presented in subject data listings.

Treatment-emergent AEs (TEAEs) will be summarized by treatment group on the Safety Analysis Set. The incidence of TEAEs will be reported as the number (percentage) of subjects with TEAEs by SOC and PT. A subject will be counted only once within a SOC and PT, even if the subject experienced more than one TEAE within a specific SOC and PT. The number (percentage) of subjects with TEAEs will also be summarized by maximum severity (mild, moderate, or severe). The number (percentage) of subjects with TEAEs will also be summarized by relationship to study drug (possibly related, probably related, and not related). TEAEs of special interest will be summarized similarly.

Adverse events will be summarized by the following subgroups: age ($\leq$ 65 years, 65 < age < 80, $\geq$ 80 years), sex (male, female), race (white, black, Asian, other), region, clinical subgroup (MCI due to AD – intermediate likelihood and mild Alzheimer's disease dementia), presence or absence of ongoing AD treatment (i.e., AChEIs and/or memantine), and *APOE ε4 carrier status* (positive, negative).

The number (percentage) of subjects with TEAEs leading to death will be summarized by MedDRA SOC and PT for each treatment group. A subject data listing of all AEs leading to death will be provided.

The number (percentage) of subjects with treatment-emergent serious adverse events (SAEs) was summarized by MedDRA SOC and PT for each treatment. A subject data listing of all SAEs was provided.

The number (percentage) of subjects with TEAEs leading to discontinuation from study drug will be summarized by MedDRA SOC and PT for each treatment. A subject data listing of all AEs leading to discontinuation from study drug will be provided.

5.6.3 C-SSRS

The C-SSRS responses will be mapped to C-CASA. The incidence of Treatment Emergent suicidal ideation or suicidal behavior in treated subjects will be summarized by treatment group. Continuous variables will be summarized by descriptive statistics; number of subjects, mean, standard deviation, median, minimum, and maximum and categorical variables by number (percentage) of subjects. Subjects who already had suicidal ideation or behavior at baseline will be excluded from summaries.

5.6.4 ANTI-DRUG (BAN2401) ANTIBODIES (ADA)

Definitions for BAN2401 Treatment Arms

ADA analysis for BAN2401 treatment arms will be based on evaluable subjects, which is defined as subjects who were treated with BAN2401 and had at least one valid post treatment ADA sample. ADA positive subjects are defined as those subjects with a treatment-induced or treatment-boosted ADA response, where treatment-induced response means ADA developed following drug administration, and treatment-boosted response means ADA already existed before treatment and boosted to a higher titer following drug administration. ADA negative subjects are those without a treatment-induced or treatment-boosted ADA response. It is possible in some instances that some subjects who did not have a measured positive ADA response during the treatment period (i.e., were negative for ADA) may not be confirmed negative due to drug interference in the sample testing process. Those subjects will be classified as ADA negative*. This distinction will be reserved only for those subjects who had PK level ≥ drug tolerance limit of the assay and did not have positive ADA samples at any visits, and did not have ADA samples collected at the 3-month follow up visit. If a subject had a negative sample at the 3-month follow up visit and negative samples at all previous visits, then the subject is considered to be confirmed ADA negative. ADA incidence rate is defined as the proportion of ADA positive subjects among all ADA evaluable subjects.

Definitions for Placebo Treatment Arm

For placebo arm, subjects are unevaluable for an immune response to study drug treatment but will be evaluated for a baseline ADA response. Placebo subjects with a baseline ADA response are defined as those with at least one positive ADA sample at any visit.

ADA prevalence rate is defined as the proportion of subjects with baseline ADA response in placebo arm plus BAN2401 subjects with positive ADA samples prior to treatment among all treated subjects.

Analysis

ADA incidence rate and the maximum titer for ADA positive subjects will be summarized by treatment group. ADA prevalence rate will also be summarized.

Clinical efficacy data will be analyzed by ADA subgroups (positive/negative). Analyses will also be performed to investigate the potential association between ADA and AEs that can be attributed to immunogenicity. For the efficacy and safety analyses, ADA negative and ADA negative* subjects will be combined into one negative group.

The efficacy and safety analyses will also be repeated for neutralizing ADAs.

More analyses for ADA/neutralizing ADA will be described in a separate ADA analysis plan document. The correlation between ADA and PK profile will also be evaluated. Details of such analyses will be described in the PK/PD analysis plan.

5.6.5 Laboratory Values

Laboratory results will be summarized using Système International (SI) units, as appropriate. For all quantitative parameters listed in protocol Section 9.5.1.5 Safety Assessments (Laboratory Measurements), the actual value and the change from baseline to each post-baseline visit and to the end of treatment (defined as the last on-treatment value) will be summarized by visit and treatment group using descriptive statistics. Qualitative parameters listed in protocol Section 9.5.1.5 will be summarized using frequencies (number and percentage of subjects), and changes from baseline to each post-baseline visit and to end of treatment will be reported using shift tables. Percentages will be based on the number of subjects with both non-missing baseline and relevant post-baseline results.

Laboratory test results will be assigned a low/normal/high (LNH) classification according to whether the value was below (L), within (N), or above (H) the laboratory parameter's reference range. Within treatment comparisons for each laboratory parameter will be based on 3-by-3 shift table that compares the baseline LNH classification to the LNH classification at each post-baseline visit and at the end of treatment. Similar shift tables will be used to compare the baseline LNH classification to the LNH classification for the highest and lowest value during the treatment period.

The Sponsor's Grading for Laboratory Values (see Appendix 13.1) presents the criteria that will be used to identify subjects with treatment-emergent markedly abnormal laboratory values (TEMAV). Except for phosphate, a TEMAV is defined as a post-baseline value with an increase from baseline to a grade of 2 or higher. For phosphate, a TEMAV is defined as a post-baseline value with an increase from baseline to a grade of 3 or higher. When displaying the incidence of TEMAVs, each subject will be counted once in the laboratory parameter high and in the laboratory parameter low categories, as applicable.

5.6.6 Vital Signs

Descriptive statistics for vital signs parameters (diastolic and systolic blood pressure, pulse, respiration rate, temperature and weight) and changes from baseline will be presented by visit and treatment group.

In addition, frequency counts of clinically notable vital signs will be summarized by treatment group. Table 3 presents the clinical notable ranges.

Table 3: Clinical Notable Ranges for Vital Signs

Vital Sign	Criterion for Low	Criterion for High
Pulse (bpm)	**< 50**	**> 100**
Temperature (°C)	**< 36**	**> 38**
Weight (kg)	**< 45**	**> 100**
Systolic BP	**< 90**	**> 160**
Diastolic BP	**< 50**	**> 100**

5.6.7 Electrocardiograms

Abnormal ECG findings will be presented as shifts from baseline (normal/abnormal clinically significant/abnormal not clinically significant) to post baseline (normal/abnormal clinically significant/abnormal not clinically significant) visits. The number and percent of subjects will be presented.

5.6.8 Other Safety Analyses

No other safety analyses are planned for this study.

5.7 OTHER ANALYSES

Other analyses may be conducted as ad-hoc to investigate impact of intercurrent events (e.g., early discontinuation and initiation of new AChEIs or memantine) on study outcomes. The reason and timing of intercurrent events, data collected after intercurrent events, missing data

mechanisms and different statistical methods for handling missing data may be explored as appropriate. Impact of potential correlation or interaction between baseline status and disease course and between subject subgroups and treatment groups on the study outcomes may be explored as appropriate. Any ad-hoc analyses that are performed will be appropriately titled/labeled as ad-hoc and will be clearly distinguished from planned analyses if results are reported in the Clinical Study Report.

5.8 EXPLORATORY ANALYSES

Exploratory Analyses will be performed on the Full Analysis Set.

Exploratory Analyses for Clinical Endpoints

Change from baseline in ADCOMS, ADAS-Cog, CDR-SB, MMSE, and FAQ across all clinical assessment time points not included in primary, key secondary or secondary analysis will be summarized by visits. The MMRM described in Section 5.4.1.2 will be used to analyze change from baseline in ADCOMS, ADAS-Cog, CDR-SB, MMSE, and FAQ across all clinical assessment time points not included in primary, key secondary or secondary analysis. The analysis will be performed with/without censoring as described in Section 5.4.1.

Exploratory Analyses for Biomarkers

Scatter plots including the Spearman correlation coefficient will be produced for change from baseline in vMRI values (total hippocampal volume, right and left hippocampal, whole brain, and total ventricular volumes) and change from baseline in amyloid PET values (SUVR of global cortical average with reference region of subcortical white matter, whole cerebellum, whole cerebellum mask, whole cerebellum adjusted by subcortical white matter, cerebellar grey matter, and composite reference consisting of cerebellar cortex, pons, subcortical white matter and cerebellar white matter) at 12 and 18 months.

Change from baseline at 12 and 18 months in PET SUVR, volumetric MRI and CSF biomarkers will also be analyzed by baseline PET SUVr subgroups (≥ median and < median). The median here is the median baseline value among all subjects.

Change from baseline in brain amyloid levels as measured by amyloid PET SUVR of global cortical average with other reference regions at 12 and 18 months will also be explored. Some other PET data parameters may also be analyzed. The amyloid PET SUVR of global cortical average with whole cerebellum as reference region and subcortical white matter as longitudinal adjustment factor can be derived as follows (Landau et al, 2015):

1) For each visit, the global cortical average (*Ctx*) is first normalized to the whole cerebellum (*WhCereb*) at the same visit:

   Baseline: $Ctx^{*}_{BL} = \frac{Ctx_{BL}}{WhCereb_{BL}}$

   At post baseline visit v_i: $Ctx^{*}_{v_i} = \frac{Ctx_{v_i}}{WhCereb_{v_i}}$

2) For each visit, calculate the ratio below for subcortical white matter (*WM*):

Baseline: $WM^*_{BL} = \frac{WM_{BL}}{WhCereb_{BL}}$

At post baseline visit v_i : $WM^*_{v_i} = \frac{WM_{v_i}}{WhCereb_{v_i}}$

3) The normalized global cortical average in 1) is then scaled by the ratio of subcortical white matter in 2).

Baseline: $SUVR_{BL} = \frac{Ctx^*_{BL}}{WM^*_{BL} / WM^*_{BL}} = Ctx^*_{BL}$

At post baseline visit v_i : $SUVR_{v_i} = \frac{Ctx^*_{v_i}}{WM^*_{v_i} / WM^*_{BL}}$

4) Finally, the change from baseline at post baseline visit v_i is:

$$\Delta SUVR = SUVR_{v_i} - SUVR_{BL}$$

After some simple algebra, it can be shown that the $SUVR_{v_i}$ in 3) can be written as follows:

Baseline: $SUVR_{BL} = Ctx^*_{BL} = \frac{Ctx_{BL}}{WM_{BL}} \cdot \frac{WM_{BL}}{WhCereb_{BL}}$

At post baseline visit v_i : $SUVR_{v_i} = \frac{Ctx^*_{v_i}}{WM^*_{v_i} / WM^*_{BL}} = \frac{Ctx_{v_i}}{WM_{v_i}} \cdot \frac{WM_{BL}}{WhCereb_{BL}}$

Therefore this SUVR is essentially the SUVR of global cortical average with subcortical white matter as reference region multiplied by the ratio of subcortical white matter and whole cerebellum at baseline.

Exploratory Analyses to Investigate the Relationship between Clinical Endpoints and Biomarkers

Clinical endpoints to be investigated include ADCOMS and CDR-SB. Biomarkers to be investigated include CSF (Aβ1-42, t-tau and p-tau), vMRI (total hippocampal volume, right and left hippocampal, whole brain, and total ventricular volumes), and PET values (SUVR of global cortical average versus the various reference regions).

The relationship between baseline biomarker values and change from baseline in clinical endpoints at 12 or 18 months will be investigated as follows:

- An MMRM will be fitted with dependent variable of change from baseline in clinical endpoints at 12 months and covariates of baseline clinical endpoints, treatment group, visit, randomization stratification variables (i.e., clinical subgroup [MCI due to AD, Mild Alzheimer's Disease Dementia], the presence or absence of ongoing AD treatment

[AChEIs or memantine or both], APOE ε4 carrier status [positive, negative]), region, treatment group-by-visit interaction, and baseline CSF (as continuous variables). The model will be repeated using baseline vMRI and baseline amyloid PET values to replace baseline CSF respectively.

- The above analyses will also be performed for change from baseline in clinical endpoints at 18 months and baseline values of CSF, vMRI and amyloid PET values.
- The same analyses will be done with baseline CSF, baseline vMRI and baseline amyloid PET values being categorized into two groups: ≥ median, < median. The median here is the median baseline value among all subjects.

The relationship between change from baseline in biomarker values and change from baseline in clinical endpoints at 12 or 18 months will be investigated as follows:

- A scatter plot by treatment group will be made between change from baseline in clinical endpoints at 12 months and change from baseline in each of the biomarkers at 12 months, respectively: CSF, vMRI and amyloid PET values. A Spearman correlation coefficient with its p-value will be displayed in each scatterplot.
- The same analyses will also be performed for change from baseline in clinical endpoints, CSF, vMRI and amyloid PET values at 18 months.
- A longitudinal plot of mean change in clinical endpoints over time (baseline, 12 months, 18 months) by treatment group and subgroups defined by change from baseline in CSF, vMRI and PET values (≥ median, < median) will also be produced. Summary statistics will also be provided.

The relationship between change from baseline in clinical endpoints and change from baseline in biomarkers at 18 months may also be evaluated using an ANCOVA model within treatment group as appropriate. Changes from baseline in clinical endpoints at 18 months will be the response variables and either continuous or categorical change of biomarkers will be independent variables. The ANCOVA model will also include baseline value of clinical endpoint as a covariate, randomization stratification variables as factors, and other terms as appropriate.

Proportion of subjects who have become amyloid negative during study will also be summarized by treatment groups. Summary of clinical endpoints will be presented for those subjects who remain amyloid positive and subjects who change from amyloid positive to amyloid negative by visual read at either 12 months or 18 months of treatment. Time to conversion from amyloid positive to amyloid negative may also be analyzed.

Additional analyses will also be performed whereby PET SUVR values are converted to the Centiloid scale (Klunk, et al., 2015).

Only subjects that have data in both clinical endpoints and each individual biomarker at specified timepoints will be included in these respective analyses.

Exploratory Analyses for Japanese Subjects

The primary efficacy endpoint (ADCOMS) and its components, ADAS-cog, CDR-SB and MMSE across all clinical assessment time points between BAN2401 and placebo in Japanese subjects with EAD will be evaluated using the descriptive statistics within Japanese subjects as well as compared with the overall study population.

Additional subgroup analysis in safety and baseline characteristics for Japanese subjects will be performed based on the needs of drug development in Japan.

6 INTERIM ANALYSES

An unblinded independent Interim Monitoring Committee (IMC) will provide oversight to ensure that the response adaptive randomization process and interim analyses perform as expected. An independent data analysis group will perform all of the interim analyses and will provide the results to the IMC.

The first interim analysis will be conducted when the first 196 subjects are randomized, again when 250 subjects are randomized, and again after each additional 50 subjects until 800 subjects are randomized. If the study reaches 800 randomized subjects, only these initial 800 subjects randomized will contribute to additional interim analyses that will be conducted every 3 months for a total of 9 months (i.e., 3 additional interim analyses). Each interim analysis will be conducted on clean data for the derived ADCOMS. The Bayesian analysis at 12 months and 18 months will be performed when all of the enrolled subjects (i.e., including any subjects who randomized after the initial 800) have been on study treatment for 12 months and 18 months, respectively, and their clinical efficacy measures evaluated.

6.1 STOPPING RULES FOR EFFICACY (EARLY SUCCESS)

The trial will be considered a success if either 1) early success is declared at an interim analysis or 2) the trial continues to completion and is declared a success at the 12 months Bayesian analysis. Since adaptive randomization decisions will be based on the 12 month treatment endpoint, a decision for early success will require that some subjects complete 12 months of treatment, and the early success decision cannot be made prior to the study randomizing 350 subjects. Interim monitoring for early success will begin at the 350 subject interim analysis. If there is greater than a 95% probability that the ED_{90} achieves a clinically significant difference from placebo during the accrual period, the trial will stop randomization and will be declared an early success. If the trial continues to completion, the trial will be considered a success, if there is at least an 80% probability that the ED_{90} achieves the clinically significant difference from placebo.

6.2 STOPPING RULES FOR FUTILITY

The study meets statistical futility criteria if the probability that the ED_{90} is better than the placebo by the CSD is less than 0.05 (with $\leq$300 subjects randomized) or 0.075 (with $\geq$350 subjects randomized) at 12 months of treatment. If the study meets statistical futility criteria, the sponsor will make the final decision pertaining to study futility after reviewing recommendations

by the DSMB. Based on Sponsor's final decision, if the study continues to randomize subjects, it will be considered a failed study.

6.3 ADAPTIVE RANDOMIZATION

Subjects will receive 18 months of treatment. Randomization to placebo or one of five dose arms of BAN2401 will be fixed for the first 196 subjects randomized in the study (4:2:2:2:2:2; placebo to each of the five active arms). Randomization probabilities to each arm will be updated at each IA such that the randomization probability will be increased for the placebo arm and arms that represent the potential target dose (ED_{90}), and simultaneously decreased for other active arms. At the time of Amendment 05 of protocol, the sponsor has agreed to the request from the Voluntary Harmonisation Procedure (VHP) committee that subjects who are confirmed *APOE4* positive (*APOE4* hetero- or homozygous) not be randomized to the 10 mg/kg, biweekly dose. After each interim analysis after protocol Amendment 05, the randomization probability vector will be split between *APOE4+* and *APOE4-* strata to ensure no subjects are enrolled on the 10 mg/kg biweekly dose. The overall probabilities as suggested by RAR will still be preserved. Details on the derivation of the adaptive randomization probabilities, are found in the Simulation Plan, see Appendix 13.2.

7 CHANGES IN THE PLANNED ANALYSES

This document has been updated to align with updates to study protocol.

8 DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

The baseline value for efficacy and safety will be defined as the most recent value reported just before first dosing.

- For subjects who start a new AD treatment (AChEIs and/or memantine) and were not on an AD treatment at randomization, a flag will be set to "Yes" at the visit where the AD treatment started and carry forward for all subsequent visits. For subjects who were on an AD treatment at the time of randomization and a dose adjustment to the AD treatment is performed after randomization, a flag will be set to "Yes" at the visit of the dose adjustment and carry forward for all subsequent visits.

The flag should be included in all efficacy data sets. The flag in the efficacy data sets will be used as a censoring mechanism for the primary endpoint described in Section 5.4.1.

8.1 DEFINITION OF VISIT WINDOWS

The protocol specified visits/weeks and corresponding time windows used for visit-wise analyses are presented in terms of days relative to the first dose. For subjects that are randomized, the last available assessment on or before the date of the first dose of study drug will be used as the Baseline-Day 1.

Different assessment may have different visit schedule. A general rule for determining a time window for each scheduled visit is to split the two consecutive scheduled visits in the middle point of the two consecutive visits. For example, the time window for visit X is from the middle

point between visit (X-1) and X (exclusive) to the middle point between visit X and (X+1) (inclusive). According to protocol, "assessments should take place on the first day of the study visit in the designated study week except as noted below (footnotes h, i, and j pertaining to imaging assessments, and footnote o pertaining to CSF sample collection)" (see page 96, Protocol Amendment 08). Therefore, the scheduled day corresponding to scheduled visit at week Y is ((Y-1)*7+1) days. For example, the scheduled day for scheduled visit at weeks 1, 3, and 5 are Day 1, Day 15, and Day 29, respectively and the corresponding time window for scheduled visit at weeks 1 and 3 are [Day 1, Day 8] and [Day 9, Day 22], respectively

For scheduled every 3 months post baseline efficacy assessments, the analysis visits are the same as the scheduled visits. Any window rules will not apply. This is per ITT principle and also assumes that the actual assessment dates are independent from any other factors or are random and centered at scheduled time point.

For Early Termination (ET) visit and Follow Up (FU) visit, the following rules apply for efficacy assessment:

- If the last scheduled 3-month assessment before early termination (LS), ET assessment and FU assessment are in three separate sequential 3-month windows, e.g., Window(x), Window (x+1) and Window(x+2), then the analysis visits for last scheduled 3-month assessment (LS), ET assessment and FU assessment will be x, x+1 and x+2, respectively.

- If both LS and ET are in Window(x) and FU is in Window(x+1) [note: this is the most likely situation since ET and FU are expected to be separated by 3 months], then analysis visit is x for LS (per rule 1.) and (x+1) for FU and ET will not be included in Bayesian analyses and MMRM analyses.

- If LS is in Window(x) and both ET and FU are in Window(x+1), then the analysis visit is x for LS and (x+1) for one of ET and FU that is closer to scheduled (x+1) visit, another one (ET or FU) will not be included in Bayesian analyses and MMRM analyses. [Note: this scenario happens only if ET is close to window (x) and FU is within three months from ET.]

8.2 ALGORITHMS FOR EFFICACY PARAMETERS

This section describes the algorithms and missing data handling procedure to derive the totals scores for the efficacy parameters FAQ, CDR-SB, MMSE, and ADA-Cog.

FAQ: The FAQ is composed of 10 activities. Each activity is rated as 0 = Normal, 1 = Has difficulty but does by self, 2 = Requires assistance, 3 = Dependent, 8 = Not Applicable. The Total Score is the sum of the 10 activities. If any activity is missing then the Total Score is missing. Activities marked as "Not Applicable", are not used in the computation of the Total Score. However, in order to account for "Not Applicable" activity(s), the Total Score is weighted as follows:

Total Score = Total Score x 30 / (30 minus 3 times the number of activities marked as "Not Applicable")

CDR-SB: The CDR-Sum of the Boxes (SB) will be determined by adding the individual scores from each of the six domains. If any is domain has missing data then the CDR-SB will be missing. The values in the data will need to be converted before performing the summation, as follows;

All CDR domains, except Personal Care, use (1 = 0, 2= 0.5, 3= 1, 4=2, and 5=3), for Personal Care, use (1 = 0, 2= 1, 3= 2, and 4=3).

MMSE: The MMSE is composed of 30 questions group into domains, see Table 4. For each of the MMSE domains add the correct responses. If a domain has missing data then the domain is missing. From the domains compute the six items as show in Table 4. If any domain is missing then the item is missing. The MMSE Total Score (range 0 to 30) = sum of the six items. If any item score is missing then the Total Score is missing.

Table 4: MMSE Domains and Items

Domain	Score Range	Item	Score Range
1. Orientation to Time	0 to 5	1. Orientation to Time	0 to 5
2. Orientation to Place	0 to 5	2. Orientation to Place	0 to 5
3. Registration	0 to 3	3. Registration	0 to 3
4. Attention and Calculation[a]	0 to 5	4. Attention and Calculation	0 to 5
5. Recall	0 to 3	5. Recall	0 to 3
6. Naming	0 to 2	6. Language (Sum of Naming, Repetition, Comprehension, Reading, Writing, and Drawing)	0 to 9
7. Repetition	0 to 1		
8. Comprehension	0 to 3		
9. Reading	0 to 1		
10. Writing	0 to 1		
11. Drawing	0 to 1		
		Total Score	**0 to 30**

[a] Spell WORLD Forward, then Backward score is only use if Attention and Calculation score is not available

ADAS-Cog: The ADAS-Cog is composed of 14 items, see Table 5.

Table 5: ADAS-Cog Items and Algorithm for derivation of Item Scores and Total Score

Item	Algorithm	Handling Missing Data	Score Range
1. Word Recall	Total the number of "No" responses for each trial. The subscore is the sum the scores from trials 1, 2, and 3, divide by 3.	If a trial is missing or partially done then trial score is missing. If a trial score is missing then the subscore is missing	0 to 10
2. Commands	Total the number of "No" responses from the 5 tasks	If any task is missing then the subscore is missing	0 to 5
3. Constructional Praxis	Count the number of "No" responses. The subscore is 0 = all 4 drawings correct 1= 1 figure drawn incorrectly 2= 2 figures drawn incorrectly 3= 3 figures drawn incorrectly 4= 4 figures drawn incorrectly 5= no figures drawn, scribbles, parts of forms	If any task is missing then the subscore is missing	0 to 5
4. Delayed Word-Recall	Total the number of "No" responses.	If any response is missing then subscore is missing	0 to 10
5. Naming Objects / Fingers	Total the number of "No" responses. The subscore is 0= 0-2 "no" responses 1= 3-5 "no" responses 2= 6-8 "no" responses 3= 9-11 "no" responses 4= 12-14 "no" responses 5= 15-17 "no" responses	If any response is missing then subscore is missing	0 to 5
6. Ideational Praxis	Total the number of "No" responses.	If any response is missing then subscore is missing	0 to 5
7. Orientation	Total the number of "No" responses.	If any response is missing then subscore is missing	0 to 8
8. Word Recognition	For each trial, total the number of "1" responses. If the total is 12 or less, then the trial score = total. If the total is > 12 then trial score=12. The subscore is the sum the scores from trials 1, 2, and 3, divide by 3.	If a trial is missing or partially done then trial score is missing. If a trial score is missing then the subscore is missing	0 to 12
9. Remembering Test Instructions	The subscore is 0= None, 1= Very Mild, 2= Mild, 3= Moderate, 4= Moderately Severe, 5= Severe	If the response is missing then subscore is missing	0 to 5

10. Comprehension	The subscore is 0= None, 1= Very Mild, 2= Mild, 3= Moderate, 4= Moderately Severe, 5= Severe	If the response is missing then subscore is missing	0 to 5
11. Word Finding Difficulty	The subscore is 0= None, 1= Very Mild, 2= Mild, 3= Moderate, 4= Moderately Severe, 5= Severe	If the response is missing then subscore is missing	0 to 5
12. Spoken Language Ability	The subscore is 0= None, 1= Very Mild, 2= Mild, 3= Moderate, 4= Moderately Severe, 5= Severe	If the response is missing then subscore is missing	0 to 5
13. Executive Function (Maze)	The subscore is based on the total number of seconds to complete the task and/or whenever the task was stopped due to 2 errors being made, as follows; 0 = 0-30 seconds 1 = 31-60 seconds 2 = 61 – 90 seconds 3 = 91 – 120 seconds 4 = 121-239 seconds 5 = 240 seconds or at least 2 errors	If the response is missing then the subscore is missing	0 to 5
14. Number Cancellation	Adjusted Score = Total # correct targets crossed off minus Total # incorrect targets crossed off minus Total # times reminded of task. Then use Adjusted Score to determine the subscore as follows; Adjusted Score = Subscore ≥23 = 0 18-22 = 1 13-17 = 2 9-12 = 3 5-8 = 4 ≤4 = 5	If any component of the adjusted score is missing then the subscore is missing	0 to 5
Total Score	Total Score = sum of the subscores above	If any subscore is missing then Total Score is missing	0 to 90

9 PROGRAMMING SPECIFICATIONS

The rules for programming derivations and dataset specifications are provided in separate documents.

10 STATISTICAL SOFTWARE

All statistical analyses will be performed by Eisai Inc., using SAS Version 9.4. Tessella will use a validated algorithm to run the Bayesian NDLM.

11 MOCK TABLES, LISTINGS AND GRAPHS (TLGS)

The study TLG shells will be provided in a separate document, which will show the content and format of all tables, listings, and graphs in detail.

12 REFERENCES

1. West M, Harrison P.J. Bayesian Forecasting and Dynamic Models (2nd edn.) Springer: New York, 1997.
2. West M, Harrison P.J., Migon H.S. Dynamic generalized linear models and Bayesian forecasting. *JASA* 1985; 80:73-83.
3. Smith, M.K., et al. Implementation of Bayesian adaptive design in a proof of concept study. *Pharmaceut. Statist.* 2006; 5: 39-50.
4. Krams, M., et. al. Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN) An Adaptive Dose-Response Study of UK-279,276 in Acute Ischemic Stroke. *Stroke* 2003; 34:2543-2548.
5. Christopher JW, Spiegelhalter DJ, Grieve AP. Flexible Design and Efficient Implementation of Adaptive Dose-Finding Studies. *Journal of Biopharmaceut. Statist.* 2007;17: 1033-1050.
6. Berry SM, Carlin BP, Lee JJ, Müller P. Bayesian Adaptive Methods for Clinical Trials. CRC Press Chapman & Hall, 2001.
7. Fitsmaurice, G.M., Laird, N.M., and Ware, J.H. Applied Longitudinal Analysis. John Wiley & Sons, Inc., 2004.
8. Cnaan, A, Laird, N. M., and Slasor, P. 1997. Tutorial in Biostatistics. Using The General Linear Mixed Model to Analyze Unbalanced Repeated Measures and Longitudinal Data. Statistics in Medicine 1997; 16: 2349-2380.
9. Grady,J.J., and Helms, R. W. 1995. Model Selection Techniques for the Covariance Matrix for Incomplete Longitudinal Data. Statistics in Medicine 1995; 14: 1397-1416.
10. Littell, R.C., Pendergast, J., and Natarajan, R. Modelling covariance structure in the analysis of repeated measures data. 2000;19:1793-1819.
11. Little, R.J.A. and Rubin, D.B. Statistical Analysis with Missing Data, Second Edition, New York: John Wiley & Sons, 2002.
12. Schafer, J.L. Analysis of Incomplete Multivariate Data, New York: Chapman and Hall, 1997.
13. Landau, S., et. al., Measurement of Longitudinal b-Amyloid Change with 18F-Florbetapir PET and Standardized Uptake Value Ratios, Journal of Nuclear Medicine 2015; 56: 567-574
14. FDA, Draft Guidance for Industry: Early Alzheimer's Disease: Developing Drugs for Treatment, 2018.
15. European Medicines Agency, Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease, 2018
16. European Prevention of Alzheimer's Dementia (EPAD) Protocol for Trial EPAD-UoE-002, A double-blind, randomized, placebo-controlled, adaptive platform, Proof-of Concept (PoC) clinical trial of multiple interventions for the secondary prevention of Alzheimer's Dementia in subjects from preclinical to prodromal stages of Alzheimer's Disease, 2016
17. Klunk, W.E., Koeppe, R.A., Price, J.C., Benzinger, T.L., Devous, M.D. Sr, Jagust, W.J., et al. The Centiloid project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dementia. 2015;11:1-15

13 APPENDICES

13.1 Sponsor's Grading for Determining Markedly Abnormal Laboratory Results

The following table of Sponsor's Grading for Laboratory Values is copied from the protocol, Appendix 1.

Sponsor's Grading for Laboratory Values

	Grade 1	Grade 2	Grade 3	Grade 4
BLOOD/BONE MARROW				
Hemoglobin	< LLN – 10.0 g/dL < LLN – 100 g/L < LLN – 6.2 mmol/L	< 10.0 – 8.0 g/dL < 100 – 80 g/L < 6.2 – 4.9 mmol/L	< 8.0 g/dL < 80 g/L < 4.9 mmol/L; transfusion indicated	life-threatening consequences; urgent intervention indicated
Leukocytes (total WBC)	< LLN – 3.0 x 10^9/L < LLN – 3000/mm^3	< 3.0 – 2.0 x 10^9/L < 3000 – 2000/mm^3	< 2.0 – 1.0 x 10^9/L < 2000 – 1000/mm^3	< 1.0 x 10^9/L < 1000/mm^3
Lymphocytes	< LLN – 800/mm^3 < LLN – 0.8 x 10^9/L	< 800 – 500/mm^3 < 0.8 – 0.5 x 10^9/L	< 500 – 200/mm^3 < 0.5 – 0.2 x 10^9/L	< 200/mm^3 < 0.2 x 10^9/L
Neutrophils	< LLN – 1.5 x 10^9/L < LLN – 1500/mm^3	< 1.5 – 1.0 x 10^9/L < 1500 – 1000/mm^3	< 1.0 – 0.5 x 10^9/L < 1000 – 500/mm^3	< 0.5 x 10^9/L < 500/mm^3
Platelets	< LLN – 75.0 x 10^9/L < LLN – 75,000/mm^3	< 75.0 – 50.0 x 10^9/L < 75,000 – 50,000/mm^3	< 50.0 – 25.0 x 10^9/L < 50,000 – 25,000/mm^3	< 25.0 x 10^9/L < 25,000/mm^3
METABOLIC/LABORATORY				
Albumin, serum- low (hypoalbuminemia)	< LLN – 3 g/dL < LLN – 30 g/L	< 3 – 2 g/dL < 30 – 20 g/L	< 2 g/dL < 20 g/L	life-threatening consequences; urgent intervention indicated
Alkaline phosphatase	> ULN – 3.0 x ULN	> 3.0 – 5.0 x ULN	> 5.0 – 20.0 x ULN	> 20.0 x ULN
ALT	> ULN – 3.0 x ULN	> 3.0 – 5.0 x ULN	> 5.0 – 20.0 x ULN	> 20.0 x ULN
AST	> ULN – 3.0 x ULN	> 3.0 – 5.0 x ULN	> 5.0 – 20.0 x ULN	> 20.0 x ULN
Bicarbonate, serum-low	< LLN – 16 mmol/L	< 16 – 11 mmol/L	<11 – 8 mmol/L	< 8 mmol/L
Bilirubin (hyperbilirubinemia)	> ULN – 1.5 x ULN	> 1.5 – 3.0 x ULN	> 3.0 – 10.0 x ULN	> 10.0 x ULN
Calcium, serum-low (hypocalcemia)	< LLN – 8.0 mg/dL < LLN – 2.0 mmol/L	< 8.0 – 7.0 mg/dL < 2.0 – 1.75 mmol/L	< 7.0 – 6.0 mg/dL < 1.75 – 1.5 mmol/L	< 6.0 mg/dL < 1.5 mmol/L
Calcium, serum-high (hypercalcemia)	> ULN – 11.5 mg/dL > ULN – 2.9 mmol/L	> 11.5 – 12.5 mg/dL > 2.9 – 3.1 mmol/L	> 12.5 – 13.5 mg/dL > 3.1 – 3.4 mmol/L	> 13.5 mg/dL > 3.4 mmol/L
Cholesterol, serum-high (hypercholesterolemia)	> ULN – 300 mg/dL > ULN – 7.75 mmol/L	> 300 – 400 mg/dL > 7.75 – 10.34 mmol/L	> 400 – 500 mg/dL > 10.34 – 12.92 mmol/L	> 500 mg/dL > 12.92 mmol/L
Creatinine	> ULN – 1.5 x ULN	> 1.5 – 3.0 x ULN	> 3.0 – 6.0 x ULN	> 6.0 x ULN
GGT (γ-Glutamyl transpeptidase)	> ULN – 3.0 x ULN	> 3.0 – 5.0 x ULN	> 5.0 – 20.0 x ULN	> 20.0 x ULN
Glucose, serum-high (hyperglycemia)	Fasting glucose value: > ULN – 160 mg/dL > ULN – 8.9 mmol/L	Fasting glucose value: > 160 – 250 mg/dL > 8.9 – 13.9 mmol/L	> 250 – 500 mg/dL; > 13.9 – 27.8 mmol/L; hospitalization indicated	> 500 mg/dL; > 27.8 mmol/L; life-threatening consequences
Glucose, serum-low (hypoglycemia)	< LLN – 55 mg/dL < LLN – 3.0 mmol/L	< 55 – 40 mg/dL < 3.0 – 2.2 mmol/L	< 40 – 30 mg/dL < 2.2 – 1.7 mmol/L	< 30 mg/dL < 1.7 mmol/L life-threatening consequences; seizures
Phosphate, serum-low (hypophosphatemia)	< LLN – 2.5 mg/dL < LLN – 0.8 mmol/L	< 2.5 – 2.0 mg/dL < 0.8 – 0.6 mmol/L	< 2.0 – 1.0 mg/dL < 0.6 – 0.3 mmol/L	< 1.0 mg/dL < 0.3 mmol/L

Sponsor's Grading for Laboratory Values

	Grade 1	Grade 2	Grade 3	Grade 4
				life-threatening consequences
Potassium, serum-high (hyperkalemia)	> ULN – 5.5 mmol/L	> 5.5 – 6.0 mmol/L	> 6.0 – 7.0 mmol/L hospitalization indicated	> 7.0 mmol/L life-threatening consequences
Potassium, serum-low (hypokalemia)	< LLN – 3.0 mmol/L	< LLN – 3.0 mmol/L; symptomatic; intervention indicated	< 3.0 – 2.5 mmol/L hospitalization indicated	< 2.5 mmol/L life-threatening consequences
Sodium, serum-high (hypernatremia)	> ULN – 150 mmol/L	> 150 – 155 mmol/L	> 155 – 160 mmol/L hospitalization indicated	> 160 mmol/L life-threatening consequences
Sodium, serum-low (hyponatremia)	< LLN – 130 mmol/L	N/A	< 130 – 120 mmol/L	< 120 mmol/L life-threatening consequences
Triglyceride, serum-high (hypertriglyceridemia)	150 – 300 mg/dL 1.71 – 3.42 mmol/L	> 300 – 500 mg/dL > 3.42 – 5.7 mmol/L	> 500 – 1000 mg/dL >5.7 – 11.4 mmol/L	> 1000 mg/dL > 11.4 mmol/L life-threatening consequences
Uric acid, serum-high (hyperuricemia)	> ULN – 10 mg/dL ≤ 0.59 mmol/L without physiologic consequences	N/A	> ULN – 10 mg/dL ≤ 0.59 mmol/L with physiologic consequences	> 10 mg/dL > 0.59 mmol/L life-threatening consequences

ALT = alanine aminotransferase (serum glutamic pyruvic transaminase), AST = aspartate aminotransferase (serum glutamic oxaloacetic transaminase), GGT = γ-glutamyl transpeptidase, N/A = not applicable, LLN = lower limit of normal, ULN = upper limit of normal, WBC = white blood cell.
Based on Common Terminology Criteria for Adverse events (CTCAE) Version 4.0. Published: May 28, 2009 (v4.03: June 14, 2010).

13.2 Study Simulation Plan

Appendix 13.2 presents the simulation plan and details the Normal Dynamic Linear Model, simulations, design, and operating characteristics.

APPENDIX 13.2 Simulation Plan

Phase II Dose Finding Study of BAN2401 in Patients with Mild Cognitive Impairment Due to Alzheimer's Disease and Mild Alzheimer's Disease

Submitted to Eisai
June 25, 2012

13.3 **TABLE OF CONTENTS**

1.0 INTRODUCTION

This is a phase II dose-finding study of BAN2401 for the treatment of Alzheimer's disease. The minimum sample size will be 196 patients and the maximum sample size will be 800 patients. We use response adaptive randomization to allocate patients between placebo control and 5 active doses with the goal of characterizing the dose response curve. The study will be monitored for early success and early futility. There are two schedules of administration, bi-weekly and once monthly. The primary endpoint is the change in a composite clinical score from baseline to 52 weeks. A difference from control of at least 0.03 would be considered a clinically significant difference (CSD).

2.0 STATISTICAL MODELING

2.1 Primary Endpoint Modeling

The primary endpoint is the change from baseline to 52 weeks in a composite clinical score. We label the observations of the change from baseline in composite clinical score for subject *i*, at the 6, 12, 26, 39, and 52 week visits as $Y_{i,6}$, $Y_{i,12}$, $Y_{i,26}$, $Y_{i,39}$, and $Y_{i,52}$, respectively. We model the 52-week primary outcomes as normally distributed,

$$[Y_{i,52}] \sim N(\theta_d, \sigma^2),$$

where d_i is the treatment arm for subject *i*. The treatment arms are: d=0 (control), *d*=1 (3 mg/kg bi-weekly), *d*=2 (5 mg/kg bi-weekly), *d*=3 (10 mg/kg bi-weekly), *d*=4 (5 mg/kg monthly), and *d*=5 (10 mg/kg monthly).

2.2 Dose-Response Model Primary

We label the mean change from baseline to 52-weeks for dose group *d* as θ_d. We construct a dose-response model for the mean change from baseline for each treatment arm. The arms are modeled with a two-dimensional first-order normal dynamic linear model (NDLM). This model is a Gaussian random walk model. The structure is

$$[\theta_0] \sim N(0, 0.5^2)$$
$$[\theta_1] \sim N(0, 0.5^2)$$

and the NDLM structure for the bi-monthly doses:

$$[\theta_2] \sim N(\theta_1, \tau^2)$$

$$[\theta_3] \sim N(\theta_2, \tau^2)$$

The monthly doses are "connected" to the dose-response model through the respective bi-monthly doses. The connection through the NDLM first order structure results in the following priors

$$[\theta_4] \sim N(\theta_2, \tau^2)$$

and

$$[\theta_5] \sim N\left(\frac{\theta_3 + \theta_4}{2}, \frac{\tau^2}{2}\right).$$

The drift parameter (variance component), τ^2 is modeled with the following prior distributions:

$$[\tau^2] \sim IG(0.25, 0.0025)$$

where *IG(a,b)* is the inverse gamma distribution with shape parameter *a* and scale parameter *b*. This prior has the equivalent weight of 0.5 observations worth of weight that the value of τ is 0.1.

The prior distribution on the initial dose is selected to be a vague prior. An effect size of 0.03 is considered clinically significant and the standard error of the clinical scale is 0.15. Therefore these priors are quite weak and allow the data to shape the dose-response model. The prior for the "drift parameter" is selected to have minor weight that there will be similarity across the doses, and hence a minor amount of borrowing. With 5 doses the amount of borrowing is determined largely by the similarity in outcomes across doses.

The distribution of the variance of the primary endpoint, σ^2, is modeled with the following prior:

$$[\sigma^2] \sim IG(2.5, 0.056).$$

The prior is equivalent to a prior weight of 5 observations worth of information that the value of σ is 0.15. The value of 0.15 is based on initial modeling by Eisai. This prior allows the historical data estimate of 0.15 to have some role in the early part of the trial. When 300 subjects have been enrolled this prior will have a very minor role (5/305 weight is less than 2%).

2.3 Longitudinal Model

At each interim analysis there will be subjects who have complete or incomplete information. Some subjects will have complete information meaning they have a 52-week value observed, $Y_{i,52}$. These subjects may also have their earlier values observed, $Y_{i,6}$, $Y_{i,12}$, $Y_{i,26}$, and $Y_{i,39}$. There will be subjects with earlier observations, but no 52-week observation. There will be subjects with no observations.

We utilize the information from subjects with incomplete information to the extent that the earlier values are correlated to the final 52-week value. A Bayesian model is built to learn from the accruing information on the association between the early values and the final endpoint at 52 weeks. Different sets of parameters (instances) are used for each arm in the trial, but the models correlating early to late outcomes are the same.

A linear regression model is created for the correlation between the 6-, 12-, 26-, and 39-week values and the 52-week value (we suppress the subject index *i* and refer to the change from baseline at weeks 6, 12, 26, 39, and 52 as Y_6, Y_{12}, Y_{26}, Y_{39}, and Z_{52}). A common model is used for each active dose and a separate model for the control. Each of the model instances is identical, with the following structure for early time period j:

$$\begin{aligned}
\left[Z_{52} \mid Y_j\right] &\sim N\left(\alpha_j + \beta_j Y_j, \lambda_j^2\right) \\
\left[\alpha_j\right] &\sim N\left(0, 0.05^2\right) \\
\left[\beta_j\right] &\sim N\left(0.80, 0.25^2\right) \\
\left[\lambda_j^2\right] &\sim IG(2.5, 0.025).
\end{aligned}$$

The prior distribution for the standard deviation about the linear regression, λ, has a prior mean estimate of 0.10 with a weight of 5 observations. The joint posterior distribution of α_j, β_j, and λ_j, for j=6, 12, 26, and 39, are updated at each interim analysis (as described in Section 4.0) with the observed values of Y_{52}.

These prior distributions were selected from historical data provided by Eisai. The data were discounted to allow the new study to shape the posteriors, but these priors provide empirically based starting values.

At each interim analysis (as described in Section 4.0), this model is used, with Bayesian imputation within Markov chain Monte Carlo, to update the dose-response model at each interim analysis. The Markov chain Monte Carlo steps are as follows:

1) .Simulate longitudinal model parameters from the updated joint posterior distribution
2). For patients that do not yet have a 52 week observation, impute the 52 week observation based on their most recent observation (6, 12, 26, or 39 weeks)
3). Simulate dose-response model parameters for each arm based on the "complete" (observed and imputed) 52-week data.
4). Repeat

3.0 POSTERIOR QUANTITIES

3.1 Definition of Doses

We characterize the dose-response curve by defining the maximum effective dose (d_{Max}) and the effective dose 90 (d_{ED90}). The maximum effective dose (d_{Max}) is the dose with the greatest

treatment effect (greatest difference from control in mean change from baseline to 52 weeks). The d_{ED90} is the simplest dose that achieves at least 90% of the treatment effect achieved by d_{Max}. The simplest dose level is the smallest dose level with the lowest frequency of administration. The dose levels ordered by frequency of administration and amount are 5QM, 10QM, 3BM, 5BM, and 10BM. The doses identified as the d_{Max} and the d_{ED90} may be different doses or the same dose. For example, if only one dose is effective, it would be both the d_{Max} and d_{ED90}.

We estimate the probability each dose is d_{Max}, $\Pr(d = d_{Max})$, and the probability each dose is the d_{ED90}, $\Pr(d = d_{ED90})$. Based on the posterior distributions of the parameters in the dose response model, we sample 10,000 dose-response curves. We identify the d_{Max} and the d_{ED90} for each of these curves. The proportion of simulated dose-response curves where $d = d_{Max}$ is the $\Pr(d = d_{Max})$ for a given dose d. Similarly, the proportion of simulated dose-response curves where $d = d_{ED90}$ is the $\Pr(d = d_{ED90})$ for a given dose d.

3.2 Probability Superior to Control

For each dose we calculate the probability of being superior to control by comparing the posterior distribution of the mean change from baseline to 52-weeks between the active dose and the control.

We determine the probability of being superior to control by randomly drawing 10,000 treatment effects from the posterior distribution of θ_d for the active dose, and a matching 10,000 samples from the posterior distribution of θ_0 (treatment effects for the control). Thus, we have 10,000 pairs of active dose and control treatment effects. The proportion of pairs where the treatment effect for the active dose is greater than for control is the probability of being superior to control, $\Pr(\theta_d > \theta_0)$.

3.3 Probability Superior to Control by Clinically Significant Difference

Similar to above in Section 3.2, we calculate the probability of being superior to control by at least the clinically significant difference (CSD) of 0.03 by comparing the posterior distribution of the mean change from baseline to 52 weeks between the active dose and the control.

We determine the probability of being superior to control by at the least the CSD by sampling 10,000 treatment effects from the posterior distribution of θ_d for the active dose, and a matching 10,000 samples from the posterior distribution of θ_0 (treatment effect for the control). Thus, we have 10,000 pairs of active dose and control treatment effects. The proportion of pairs where the treatment effect for the active dose is greater than for control by at least 0.03 is the probability of being superior to control by at least the CSD, $\Pr(\theta_d > \theta_0 + 0.03)$.

3.4 Probability of Phase III Success

For each active dose, we calculate the predictive probability of success in a phase III trial. We assume a hypothetical future phase III trial with a fixed design that would equally randomize 500 patients between control and an active dose (250 per arm). The predictive probability of phase III success is the chance of statistical significance ($\Pr(\theta_0 > \theta_d) < 0.025$) versus control if the active dose was entered into the hypothetical future phase III trial.

This is different from the power for such a trial, in that power calculations typically assume a fixed treatment effect whereas the predictive probability of success averages the power over the posterior distribution of the treatment effect. Thus knowledge of the treatment effect and the uncertainty in that knowledge are formally incorporated.

We determine the predictive probability of phase III success for a dose, *d*, by randomly sampling one value from the posterior distribution of the treatment effect for that dose and a corresponding random sample from the posterior distribution of the treatment effect for control. Assuming these treatment effects, we calculate the corresponding power (i.e., the probability of phase III success or the probability that $\Pr(\theta_0 > \theta_d) < 0.025$ given the randomly sampled treatment effects) and record that result. We repeat this 10,000 times. The average of the resulting probabilities of phase III success is the predictive probability of phase III success.

4.0 ALLOCATION

There will be an initial burn-in period in which patients will be allocated in blocks of 14 with 4:2:2:2:2:2 blocked to control and each of the 5 active dose arms respectively. There will be a total of 14 blocks and so a total of 196 patients will be randomized with 56 on the control and 28 on each of the active doses.

After this initial burn-in, adaptive randomization will begin. Adaptive randomization probabilities will be updated when 196, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, and 750 patients have been enrolled. The randomization probability for each of the 5 active doses will be weighted according to the variance components,

$$V_d = \Pr\left(d = d_{ED90}\right)\sqrt{\frac{Var\left(\theta_d\right)}{n_d + 1}} \text{ for d = 1, 2, 3, 4, 5}$$

where $\Pr\left(d = d_{ED90}\right)$ is the probability the dose is the ED_{90}, $Var\left(\theta_d\right)$ is the posterior variance of the treatment effect, and n_d is the current number of subjects allocated to dose *d*.

The randomization probability to the control is meant to mirror the randomization probability to the most likely ED_{90} dose. The randomization probability for the control is weighted according to

$$V_0 = \min\left\{\sum_{d=1}^{5} V_d\left(\frac{n_d}{n_d + n_0}\right), \max\left(V_1, V_2, V_3, V_4, V_5\right)\right\} \text{ for d = 0}$$

The randomization vector is then

$$q_d = \frac{V_d}{\sum_{j=0}^{5} V_j} \text{ for d = 0,1,2,3,4,5}$$

5.0 Evaluation of Trial Success and Futility

Interim analyses are planned beginning when 196 patients are enrolled and after approximately every additional 50 patients are enrolled (after 250, 300, 350, etc) as described in Section 4.0. If the trial should continue to the maximum sample size of 800 patients, post accrual interim analyses are planned 3, 6, and 9 months after the 800th patient is enrolled. If the trial is not stopped early for futility, all enrolled patients will be followed for 18 months. The final evaluation of efficacy will occur when all enrolled patients have completed follow-up for the primary endpoint.

5.1 Early Futility

Interim monitoring for early futility will occur at each interim analysis during and post accrual. Interim monitoring for futility will be based on the dose identified as the ED_{90} (d_{ED90}). If, at the first three interim looks (when 196, 250, and 300 patients have been enrolled), there is less than a 5% probability that the d_{ED90} achieves a clinically significant difference from control, the trial will stop early for futility. Formally if:

$$\Pr(\theta_{dED90} > \theta_0 + 0.03) < 0.05.$$

Beginning at the 350 patient look and continuing to the completion of the trial (complete enrollment and complete follow-up), the futility criteria is increased to 7.5%. Thus, it becomes easier to stop for futility once at least 350 patients have been enrolled. Formally the trial will stop for futility at these later interim analyses if,

$$\Pr(\theta_{dED90} > \theta_0 + 0.03) < 0.075.$$

5.2 Early Success

Interim monitoring for early success will occur at each interim analysis beginning when 350 patients have been enrolled. If there is greater than a 95% probability that the d_{ED90} achieves a clinically significant difference from control, the trial will stop for early success. Formally if:

$$\Pr(\theta_{dED90} > \theta_0 + 0.03) > 0.95.$$

5.3 Trial Completion

If the trial is not stopped early, either during accrual or post accrual, for either futility or success, then trial success will be evaluated at the completion of the trial, when both accrual and follow-up for the primary endpoint are complete. If, at the completion of the trial, there is greater than an 80% probability that the d_{ED90} achieves a clinically significant difference from control, this trial will be considered a success. Formally,

$$\Pr(\theta_{dED90} > \theta_0 + 0.03) > 0.80$$

5.4 Consideration of Phase III

At the conclusion of the study, the d_{ED90} will be considered as promising for phase III if the predictive probability of success in a future phase III trial is at least 80%.

5.5 Superior to Control

At the conclusion of the study, the d_{ED90} will be considered superior to control if the posterior probability the d_{ED90} has greater mean change at 52-weeks is at least 97.5%.

6.0 ACCRUAL AND DROPOUT

We assume accrual will increase at a constant rate over the first 6 months of the study to a peak accrual rate of 32 patients per month. The table below shows the approximate accrual rate at each month during the 6 month ramp-up to the steady state accrual rate of 32 patients per month.

Accrual Ramp-Up: Patients per Month	
Week	**Patients Per Month**
1	5.33
2	10.66
3	16
4	21.33
5	26.66
6	32
Steady State Patients per Month from 6 Months	32

We assume an exponential rate of dropout such that 22% of patients have dropped out by 26 weeks and 40% have dropped out by 52 weeks. As a sensitivity analysis we also consider dropout rates of 20% and 30% by 52 weeks.

7.0 EXAMPLE TRIAL

In this section we present an example trial to illustrate the adaptive design.

The first interim analysis occurs when 196 patients have been enrolled in a 4:2:2:2:2:2 ratio to control and each of the active doses. Table 7.1.A shows the observed data at this interim analysis. Some patients are newly enrolled or have no follow-up data yet. The longest follow-up of any patient is 26 weeks. The 5Q, 10Q, and 3BM doses have the greatest observed mean changes at 26 weeks, but the 10Q dose has a worse result at 12 weeks. Thus, the 5Q and 3BM doses have similar probability of being the ED_{90} (Table 7.1.B). The 5Q dose, the most likely ED_{90} has a 38% probability of being better than control by the clinically significant difference (CSD). The early futility stopping boundary is not met and so accrual to the trial continues with the 5Q and 3BM doses receiving a higher probability of randomization for the next 50 patients to be enrolled.

Table 7.1.A: Observed Data at the Interim Analysis with 196 Pts Enrolled												
		No Data	6 Weeks		12 Weeks		26 Weeks		39 Weeks		52 Weeks	
	N	N	N	Obs	N	Obs	N	Obs	N	Obs	N	Obs
PBO	56	12	15	0.01	26	-0.002	3	-0.085	0	0	0	0
5Q	28	7	7	-0.019	10	0.028	4	0.04	0	0	0	0
10Q	28	7	8	0.005	11	-0.02	2	0.083	0	0	0	0
3BM	28	10	7	-0.003	10	0.016	1	0.114	0	0	0	0
5BM	28	10	6	0.008	9	0.024	3	0.005	0	0	0	0
10BM	28	7	10	0.008	8	0.023	3	-0.005	0	0	0	0

Table 7.1.B: Model Estimates				
	Model Est	Pr(ED_{90})	Pr(PBO)	Pr(CSD)
PBO	-0.003	-	-	-
5Q	0.012	0.251	0.605	0.377
10Q	0.002	0.149	0.533	0.309
3BM	0.011	0.247	0.589	0.375
5BM	0.013	0.177	0.617	0.378
10BM	0.011	0.176	0.606	0.379

At the next interim analysis when 250 patients have been enrolled (Table 7.2.A) most patients still have 26 weeks or less of follow-up. The observed data is slightly worse for the 3BM dose. The 5Q, 10Q, and 3BM doses all have similar probability of being the ED_{90} (Table 7.2.B). Again, the early futility stopping boundary is not met and so accrual to the trial continues with the 5Q, 10Q, and 3BM doses receiving a similar probability of randomization for the next 50 patients to be enrolled.

Table 7.2.A: Observed Data at the Interim Analysis with 250 Pts Enrolled												
		No Data	6 Weeks		12 Weeks		26 Weeks		39 Weeks		52 Weeks	
	N	N	N	Obs	N	Obs	N	Obs	N	Obs	N	Obs
PBO	67	13	15	-0.001	31	0.005	8	-0.047	0	0	0	0
5Q	38	13	8	-0.006	12	0.016	5	0.031	0	0	0	0
10Q	35	9	8	0.00	12	-0.006	6	0.005	0	0	0	0
3BM	39	15	8	0.001	12	0.008	3	0.011	1	0.244	0	0
5BM	35	11	7	-0.005	13	0.018	4	0.006	0	0	0	0
10BM	36	10	8	0.006	13	0.01	5	0.02	0	0	0	0

Table 7.2.B: Model Estimates				
	Model Est	**Pr(ED_{90})**	**Pr(PBO)**	**Pr(CSD)**
PBO	-0.004	-	-	-
5Q	0.007	0.218	0.586	0.341
10Q	0.004	0.201	0.56	0.314
3BM	0.01	0.252	0.602	0.381
5BM	0.008	0.161	0.585	0.348
10BM	0.009	0.168	0.593	0.355

Interim analyses are conducted again after 300 and 350 patients are enrolled and results are similar. The possibility for early success stopping begins at the 350 patient interim analysis. At the interim analysis when 400 patients have been enrolled, patients are beginning to complete the 52 weeks of follow-up for the primary endpoint (Table 7.3.A). Each of the BM doses have a positive observed effect at 39 weeks, but the 3BM dose also has a strong effect at 12 weeks and a patient with a large effect at 52 weeks. The 3BM dose is the most likely ED_{90} (Table 7.3.B). The probability it is better than control by the CSD is 61%. Neither the early futility nor early success criteria are met. Accrual continues with the 3BM dose receiving a higher probability of randomization for the next 50 patients to be enrolled.

Table 7.3.A: Observed Data at the Interim Analysis with 400 Pts Enrolled												
		No Data	**6 Weeks**		**12 Weeks**		**26 Weeks**		**39 Weeks**		**52 Weeks**	
	N	**N**	**N**	**Obs**	**N**	**Obs**	**N**	**Obs**	**N**	**Obs**	**N**	**Obs**
PBO	101	14	17	0.005	34	0.005	27	-0.012	9	-0.043	0	0
5Q	63	12	11	-0.008	24	0.004	11	-0.001	4	-0.014	1	-0.001
10Q	55	10	8	-0.005	18	0.016	12	-0.002	6	-0.029	1	0.208
3BM	73	17	17	0.006	27	0.027	9	0.007	2	0.036	1	0.174
5BM	53	11	5	0.006	18	0.014	14	0.011	3	0.095	2	-0.056
10BM	55	7	11	0.006	17	0.007	13	0.012	6	0.086	1	0.085

Table 7.3.B: Model Estimates				
	Model Est	**Pr(ED_{90})**	**Pr(PBO)**	**Pr(CSD)**
PBO	-0.005	-	-	-
5Q	0.001	0.07	0.57	0.284
10Q	0.024	0.277	0.744	0.502
3BM	0.037	0.414	0.838	0.606
5BM	0.003	0.042	0.568	0.281
10BM	0.026	0.196	0.761	0.507

Results are similar at the next interim analysis when 450 patients are enrolled. At the interim analysis with 500 patients enrolled, the 10BM dose now has the largest effect and the

randomization probabilities will shift such that the 10BM dose will receive a greater weight (Tables 7.4.A and 7.4.B).

Table 7.4.A: Observed Data at the Interim Analysis with 500 Pts Enrolled

		No Data	6 Weeks		12 Weeks		26 Weeks		39 Weeks		52 Weeks	
	N	N	N	Obs	N	Obs	N	Obs	N	Obs	N	Obs
PBO	129	17	16	0	38	0.012	26	-0.005	24	-0.001	8	0.016
5Q	66	6	4	-0.015	24	0.005	18	0	9	0.009	5	-0.015
10Q	69	8	10	-0.004	18	0.005	17	0.044	10	0.036	6	-0.021
3BM	111	27	16	-0.006	40	0.018	17	0.052	8	0.026	3	0.099
5BM	55	6	0	0	15	0.026	19	0.024	10	0.04	5	0.042
10BM	70	9	10	0.008	15	0.02	20	0.039	10	0.007	6	0.119

Table 7.4.B: Model Estimates

	Model Est	Pr(ED_{90})	Pr(PBO)	Pr(CSD)
PBO	0.017	-	-	-
5Q	-0.001	0.03	0.3	0.09
10Q	0.01	0.028	0.415	0.104
3BM	0.035	0.206	0.692	0.355
5BM	0.033	0.115	0.698	0.331
10BM	0.066	0.621	0.934	0.729

Interim analyses are conducted with 550, 600, and 650 patients are enrolled and results are similar. At the interim analysis with 700 patients enrolled, the 10BM dose has the largest sample size and remains the most likely ED_{90} (Tables 7.5.A and 7.5.B). The estimated mean change from baseline to 52 weeks is 0.006 for control and 0.084 for the 10BM dose. The probability the 10BM dose is better than control by at least the clinically significant difference is 0.957,which is greater than the 0.95 required to stop for early success. Accrual to the trial stops, but all currently enrolled patients will continue to be followed.

Table 7.5.A: Observed Data at the Interim Analysis with 700 Pts Enrolled												
		No Data	6 Weeks		12 Weeks		26 Weeks		39 Weeks		52 Weeks	
	N	N	N	Obs	N	Obs	N	Obs	N	Obs	N	Obs
PBO	185	21	26	0.001	45	0.013	29	0.001	26	0.035	38	-0.006
5Q	72	6	4	-0.023	13	0.004	15	0.016	14	0	20	-0.022
10Q	80	9	6	0.006	10	-0.007	16	0.031	21	0.042	18	0.023
3BM	134	16	11	-0.004	38	0.004	34	0.026	20	0.015	15	0.093
5BM	69	8	4	0.014	13	0.002	12	0.037	14	0.052	18	0.027
10BM	160	30	34	0.009	44	0.011	16	0.028	17	0.054	19	0.084

Table 7.5.B: Model Estimates				
	Model Est	Pr(ED_{90})	Pr(PBO)	Pr(CSD)
PBO	0.006	0	0	0
5Q	-0.007	0.002	0.313	0.059
10Q	0.026	0.034	0.803	0.357
3BM	0.049	0.221	0.941	0.702
5BM	0.031	0.034	0.842	0.419
10BM	0.074	0.709	0.996	0.957

The final analysis is conducted when all enrolled patients have completed follow-up. Data is not complete on all patients due to dropout (Table 7.6.A). The observed change from baseline to 52 weeks is 0.014 for control and 0.068 for the 10BM dose. The model estimates are 0.012 for control and 0.067 for the 10BM dose (Table 7.6.B). The probability this dose is better than control by the clinically significant difference is 92%.

Table 7.6.A: Observed Data at the Final Analysis												
		No Data	6 Weeks		12 Weeks		26 Weeks		39 Weeks		52 Weeks	
	N	N	N	Obs	N	Obs	N	Obs	N	Obs	N	Obs
PBO	185	7	12	-0.002	22	0.009	16	0.011	14	-0.037	114	0.014
5Q	72	5	2	-0.029	11	0.005	8	0.006	2	0.212	44	0.005
10Q	80	6	3	-0.023	7	0.005	7	0.005	4	0.039	53	0.037
3BM	134	10	9	-0.005	17	0.007	9	0.015	15	0.016	74	0.041
5BM	69	5	1	0.031	7	0.014	9	0.019	5	0.003	42	0.041
10BM	160	9	9	0.013	15	0.005	23	0.035	11	0.077	93	0.068

Table 7.6.B: Model Estimates				
	Model Est	**Pr(ED_{90})**	**Pr(PBO)**	**Pr(CSD)**
PBO	0.012	0	0	0
5Q	0.017	0.018	0.591	0.133
10Q	0.035	0.066	0.881	0.371
3BM	0.036	0.061	0.904	0.383
5BM	0.037	0.09	0.879	0.419
10BM	0.067	0.765	0.999	0.923

8.0 SIMULATION SCENARIOS

In order to characterize the performance of the trial design, we simulated the trial considering different scenarios.

8.1 Longitudinal Scenarios

The changes in the composite clinical score observed at each visit are correlated within each patient. We assume a correlation of 0.80 from visit to visit. In addition, we make assumes for the fraction of the 52 week treatment effect that will be observed at each visit. We consider three scenarios, linear, symptomatic, and late onset. Table 8.1 shows the assumed fraction of the 52 week treatment effect at each visit. The linear scenario assumes a steady growth in the treatment effect through the 52 weeks, while the symptomatic scenario assumes the treatment effect will be observed relatively quickly, and the late onset scenario assumes the treatment effect will be observed later.

Table 8.1: Longitudinal Profiles					
	Week 6	**Week 12**	**Week 26**	**Week 39**	**Week 52**
Linear	0.12	0.23	0.50	0.75	1
Symptomatic	0.11	0.93	1	1	1
Late Onset	0.12	0.12	0.20	0.50	1

8.2 Response Scenarios

We consider 13 scenarios for the treatment effect on each arm.

Table 8.2 Response Scenarios

		Control	3BM	5BM	10BM	5QM	10QM
1	Null	0.00	0.00	0.00	0.00	0.00	0.00
2	Small Effect	0.00	0.01	0.02	0.03	0.01	0.02
3	BM Good	0.00	0.02	0.04	0.06	0.02	0.04
4	Q Low	0.00	0.03	0.06	0.10	0.01	0.03
5	Both Good	0.00	0.03	0.06	0.10	0.03	0.06
6	U-Shape Good	0.00	0.06	0.08	0.06	0.05	0.06
7	U-Shape Low	0.00	0.04	0.06	0.04	0.03	0.04
8	Small Increases	0.00	0.05	0.055	0.06	0.05	0.055
9	One Works	0.00	0.00	0.00	0.06	0.00	0.00
10	All 0.04	0.00	0.04	0.04	0.04	0.04	0.04
11	All 0.06	0.00	0.06	0.06	0.06	0.06	0.06
12	All 0.08	0.00	0.08	0.08	0.08	0.08	0.08
13	Medium Effect	0.00	0.015	0.03	0.04	0.015	0.03

9.0 OPERATING CHARACTERISTICS

Table 9.1 shows the probabilities of trial success and futility and the mean total sample size for each scenario shown in Table 8.2. We present operating characteristics assuming the linear longitudinal profile and a 40% dropout rate at 52 weeks. Operating characteristics based on the late onset and symptomatic longitudinal profiles are shown in the appendix. Operating characteristics based on the linear longitudinal profile but assuming a lower (20% and 30%) dropout rate at 52 weeks are also shown in the appendix. Results are similar across the three longitudinal profiles and the three dropout rates.

We show the probability of early success, probability of early futility, the probability the trial continues to completion (complete accrual and complete follow-up) and the total probability of trial success (either early or at trial completion).

Table 9.1: Probabilities of Trial Success and Futility						
Scenario		**Mean N**	**Pr(Stop Early Success)**	**Pr(Stop Early Futility)**	**Pr(Go To Trial Completion)**	**Pr (Success)**
1	Null	683.1	0.119	0.447	0.433	0.125
2	Small Effect	697.4	0.281	0.157	0.562	0.353
3	BM Good	656.7	0.660	0.044	0.296	0.801
4	Q Low	573.1	0.962	0.019	0.019	0.978
5	Both Good	562.4	0.970	0.010	0.019	0.987
6	U-Shape Good	582.9	0.921	0.012	0.067	0.972
7	U-Shape Low	646.4	0.684	0.035	0.280	0.817
8	Small Increases	612.1	0.831	0.019	0.150	0.927
9	One Works	668.5	0.544	0.121	0.335	0.692
10	All 0.04	654.2	0.612	0.042	0.346	0.748
11	All 0.06	592.2	0.897	0.013	0.090	0.961
12	All 0.08	535.9	0.986	0.005	0.010	0.994
13	Medium Effect	683.1	0.416	0.090	0.495	0.537

We evaluate Type I error by the probability of trial success in the scenario that there is no treatment effect. In the "Null" scenario, there is a 12.5% probability of trial success. This is the one-sided Type I error rate of this trial.

We evaluate power by the probability of trial success in alternative scenarios. For example, in scenarios where all doses are better than control by at least the CSD, such as "U-Shape Good" or "Small Increases" there is greater than a 90% probability of trial success. In the "Small Effect" scenario, there is only a 35% probability of trial success, but the treatment effect for each dose is at or below the clinically significant difference.

Table 9.2 shows the mean and standard deviation of the number of patients allocated to each treatment arm. The true ED_{90} for each scenario is shown in bold italics. The true ED_{90} dose tends to receive the greatest number of patients in each scenario.

Table 9.2: Mean and Standard Deviation of the Number of Patients Allocated to Each Treatment Arm

		Mean Subjects (SD)	Mean Subj Control (SD)	Mean Subj 5Q (SD)	Mean Subj 10Q (SD)	Mean Subj 3BM (SD)	Mean Subj 5BM (SD)	Mean Subj 10BM (SD)
1	Null	683.1 (154.7)	179.1 (43.2)	***106.6 (45.2)***	103.6 (44.3)	111.2 (45.9)	88.1 (38.8)	94.1 (41.1)
2	Small Effect	697.4 (147.7)	182.2 (40.9)	100.9 (39.4)	108.5 (43.3)	103.3 (40.3)	91.7 (38.8)	***110.8 (46.7)***
3	BM Good	656.7 (149.8)	170.6 (41.4)	89.1 (32.4)	104.9 (40.5)	90.4 (31.7)	87.7 (35.7)	***114.1 (46.3)***
4	Q Low	573.1 (126.8)	148.9 (35.7)	68.1 (18.0)	76.0 (23.1)	82.0 (25.4)	82.3 (32.3)	***115.8 (44.6)***
5	Both Good	562.4 (125.2)	144.8 (34.7)	74.4 (21.5)	88.9 (31.2)	75.4 (21.3)	74.1 (26.3)	***104.7 (39.4)***
6	U-Shape Good	582.9 (136.4)	150.1 (36.9)	86.3 (29.4)	88.6 (31.3)	94.5 (31.9)	***86.2 (33.1)***	77.3 (27.2)
7	U-Shape Low	646.4 (150.5)	167.8 (41.2)	93.4 (34.8)	97.2 (37.7)	104.1 (38.2)	***98.5 (41.0)***	85.4 (32.5)
8	Small Increases	612.1 (146.1)	158.2 (39.7)	96.4 (35.6)	***95.2 (35.7)***	96.5 (34.6)	79.5 (29.7)	86.3 (32.8)
9	One Works	668.5 (150.7)	177.1 (43.3)	86.3 (32.1)	86.9 (32.3)	91.1 (33.3)	73.9 (27.9)	***153.1 (62.4)***
10	All 0.04	654.2 (152.9)	170.3 (41.8)	***105.5 (40.9)***	101.5 (40.3)	104.9 (39.3)	83.6 (33.2)	88.5 (35.4)
11	All 0.06	592.2 (140.2)	153.0 (37.9)	***97.1 (35.2)***	91.6 (33.4)	95.0 (32.1)	75.8 (27.8)	79.6 (28.5)
12	All 0.08	535.9 (114.9)	137.6 (30.8)	***88.2 (28.7)***	82.7 (26.1)	86.9 (26.2)	68.9 (21.9)	71.6 (22.0)
13	Medium Effect	683.1 (150.2)	178.1 (41.5)	96.1 (36.1)	110.6 (44.8)	98.4 (37.1)	92.0 (39.1)	***107.8 (44.9)***

Table 9.3 shows the probability each dose is identified as the ED_{90}. The true ED_{90} for each scenario is shown in bold italics.

Table 9.3: Dose-Finding - Pr(ED_{90})

		5Q	10Q	3BM	5BM	10BM
1	Null	***0.23***	0.209	0.241	0.15	0.171
2	Small Effect	0.122	0.255	0.107	0.166	***0.35***
3	BM Good	0.049	0.232	0.04	0.136	***0.542***
4	Q Low	0.001	0.009	0.01	0.082	***0.898***
5	Both Good	0.012	0.117	0.006	0.06	***0.805***
6	U-Shape Good	0.132	0.207	0.202	***0.385***	0.074
7	U-Shape Low	0.11	0.197	0.19	***0.421***	0.082
8	Small Increases	0.252	***0.274***	0.164	0.134	0.177
9	One Works	0.011	0.013	0.015	0.01	***0.951***
10	All 0.04	***0.302***	0.233	0.225	0.127	0.113
11	All 0.06	***0.344***	0.234	0.219	0.104	0.1
12	All 0.08	***0.377***	0.243	0.199	0.096	0.085
13	Medium Effect	0.083	0.288	0.082	0.18	***0.366***

Table 9.4 shows the probability of success by arm for each scenario. This is the probability each arm achieves at least a 95% probability of being better than control by the CSD during the trial, or if the trial should run to completion, at least an 80% probability of being better than control by the CSD. The true ED_{90} for each scenario is shown in bold italics.

Table 9.4: Probability of Success Early or at Trial Completion

		5Q	10Q	3BM	5BM	10BM
1	Null	***0.031***	0.025	0.03	0.017	0.022
2	Small Effect	0.048	0.082	0.044	0.052	***0.126***
3	BM Good	0.06	0.165	0.048	0.112	***0.415***
4	Q Low	0.018	0.05	0.044	0.139	***0.727***
5	Both Good	0.048	0.17	0.041	0.1	***0.627***
6	U-Shape Good	0.14	0.192	0.205	***0.316***	0.12
7	U-Shape Low	0.095	0.15	0.158	***0.316***	0.098
8	Small Increases	0.208	***0.219***	0.17	0.136	0.193
9	One Works	0.02	0.017	0.018	0.011	***0.626***
10	All 0.04	***0.19***	0.172	0.171	0.104	0.11
11	All 0.06	***0.277***	0.219	0.207	0.124	0.135
12	All 0.08	***0.285***	0.232	0.217	0.139	0.122
13	Medium Effect	0.055	0.138	0.052	0.093	***0.198***

Table 9.5 show the probability each dose will have at least an 80% predictive probability of phase III trial success. The true ED_{90} for each scenario is shown in bold italics.

Table 9.5: Predictive Probability of Phase III Success > 80%						
		5Q	**10Q**	**3BM**	**5BM**	**10BM**
1	Null	***0.011***	0.007	0.009	0.005	0.005
2	Small Effect	0.022	0.058	0.017	0.037	***0.101***
3	BM Good	0.027	0.157	0.024	0.094	***0.435***
4	Q Low	0.001	0.008	0.009	0.078	***0.888***
5	Both Good	0.011	0.113	0.006	0.057	***0.800***
6	U-Shape Good	0.126	0.194	0.194	***0.375***	0.071
7	U-Shape Low	0.075	0.138	0.141	***0.348***	0.059
8	Small Increases	0.222	***0.247***	0.146	0.12	0.161
9	One Works	0.003	0.003	0.003	0.003	***0.641***
10	All 0.04	***0.196***	0.16	0.151	0.084	0.078
11	All 0.06	***0.325***	0.219	0.207	0.098	0.095
12	All 0.08	***0.375***	0.241	0.199	0.096	0.084
13	Medium Effect	0.028	0.12	0.025	0.073	***0.182***

10.0 CONCLUSIONS

With a minimum of 196 and a maximum of 800 patients, this design has a one-sided Type I error rate of approximately 12.5% and a high probability (at least 80%) of success for many scenarios in which the treatment would be considered effective. Adaptive allocation tends to allocate the greatest number of patients to the true ED_{90} and the true ED_{90} has a high probability of being identified for consideration in a future phase III trial.

APPENDIX FOR LONGITUDINAL PROFILES

The table below summarizes the operating characteristics of the design assuming the late onset longitudinal profile. Results are similar to those assuming the linear longitudinal profile. The probability of success differs by only 1% to 2%. The delay in observing the treatment effect slightly reduces the probability of early success and slightly increases the probability of early futility.

Probabilities of Trial Success and Futility - Late Onset Longitudinal Profile						
Scenario		**Mean N**	**Pr(Stop Early Success)**	**Pr(Stop Early Futility)**	**Pr(Go To Trial Completion)**	**Pr (Success)**
1	Null	680.5	0.134	0.432	0.434	0.140
2	Small Effect	685.5	0.295	0.171	0.534	0.364
3	BM Good	652.9	0.631	0.068	0.301	0.773
4	Q Low	590.3	0.940	0.032	0.028	0.964
5	Both Good	578.8	0.950	0.025	0.026	0.971
6	U-Shape Good	594.9	0.912	0.020	0.068	0.964
7	U-Shape Low	646.5	0.661	0.052	0.286	0.796
8	Small Increases	618.5	0.810	0.032	0.158	0.911
9	One Works	668.5	0.534	0.138	0.328	0.678
10	All 0.04	653.8	0.599	0.060	0.342	0.734
11	All 0.06	602.5	0.872	0.023	0.104	0.947
12	All 0.08	550.4	0.982	0.010	0.008	0.988
13	Medium Effect	676.6	0.421	0.116	0.464	0.528

The table below summarizes the operating characteristics of the design assuming the symptomatic longitudinal profile. Results are similar to those assuming the linear longitudinal profile. Observing the full treatment effect earlier in the course of follow-up slightly increases the probability of early success in scenarios where the treatment would be considered effective and slightly increases the probability of early futility in scenarios were the treatment would be considered not effective.

Probabilities of Trial Success and Futility - Symptomatic Longitudinal Profile						
Scenario		**Mean N**	**Pr(Stop Early Success)**	**Pr(Stop Early Futility)**	**Pr(Go To Trial Completion)**	**Pr (Success)**
1	Null	701.6	0.074	0.439	0.486	0.082
2	Small Effect	715.3	0.247	0.127	0.626	0.328
3	BM Good	662.2	0.648	0.023	0.329	0.804
4	Q Low	545.4	0.981	0.003	0.016	0.995
5	Both Good	533.4	0.982	0.002	0.016	0.995
6	U-Shape Good	563.7	0.940	0.002	0.058	0.983
7	U-Shape Low	648.7	0.689	0.016	0.296	0.835
8	Small Increases	605.0	0.827	0.006	0.168	0.933
9	One Works	679.6	0.566	0.085	0.350	0.722
10	All 0.04	665.0	0.593	0.027	0.380	0.738
11	All 0.06	576.9	0.901	0.003	0.096	0.966
12	All 0.08	500.3	0.993	0.000	0.006	0.999
13	Medium Effect	703.6	0.381	0.059	0.559	0.501

APPENDIX FOR DROPOUT RATES

The two tables below summarize the operating characteristics of the design assuming lower dropout rates of 20% and 30% at 52 weeks respectively. These results assume the linear longitudinal profile. In general these results are similar to those that assume a 40% dropout rate at 52 weeks. Lower dropout results in a slightly lower mean sample size and an slightly increased probability of trial success in scenarios where the treatment would be considered effective.

Probabilities of Trial Success and Futility – 20% Dropout at 52 Weeks						
Scenario		**Mean N**	**Pr(Stop Early Success)**	**Pr(Stop Early Futility)**	**Pr(Go To Trial Completion)**	**Pr (Success)**
1	Null	681.4	0.095	0.496	0.410	0.10
2	Small Effect	689.9	0.287	0.147	0.566	0.353
3	BM Good	642.0	0.693	0.038	0.269	0.817
4	Q Low	548.5	0.975	0.016	0.009	0.982
5	Both Good	536.4	0.985	0.007	0.008	0.993
6	U-Shape Good	560.6	0.950	0.009	0.041	0.984
7	U-Shape Low	632.8	0.714	0.029	0.258	0.846
8	Small Increases	593.3	0.863	0.015	0.122	0.944
9	One Works	662.6	0.611	0.099	0.291	0.760
10	All 0.04	645.1	0.632	0.037	0.332	0.768
11	All 0.06	569.9	0.914	0.011	0.075	0.970
12	All 0.08	508.5	0.994	0.002	0.004	0.998
13	Medium Effect	678.0	0.413	0.087	0.500	0.538

Probabilities of Trial Success and Futility – 30% Dropout at 52 Weeks						
Scenario		**Mean N**	**Pr(Stop Early Success)**	**Pr(Stop Early Futility)**	**Pr(Go To Trial Completion)**	**Pr (Success)**
1	Null	687.5	0.102	0.459	0.440	0.106
2	Small Effect	695.1	0.284	0.147	0.569	0.353
3	BM Good	649.0	0.668	0.043	0.289	0.809
4	Q Low	559.3	0.970	0.018	0.013	0.981
5	Both Good	551.1	0.984	0.006	0.01	0.993
6	U-Shape Good	569.5	0.941	0.010	0.049	0.982
7	U-Shape Low	636.5	0.698	0.031	0.271	0.827
8	Small Increases	605.7	0.842	0.018	0.140	0.935
9	One Works	666.8	0.582	0.105	0.313	0.732
10	All 0.04	646.5	0.617	0.046	0.337	0.757
11	All 0.06	576.9	0.905	0.015	0.080	0.963
12	All 0.08	519.5	0.992	0.004	0.004	0.996
13	Medium Effect	681.2	0.409	0.094	0.497	0.528