An Open-label, Randomized Phase 3 Efficacy Study of ASP8273 vs Erlotinib or Gefitinib in First-line Treatment of Patients with Stage IIIB/IV Non-small Cell Lung Cancer Tumors with EGFR Activating Mutations

ISN/Protocol 8273-CL-0302

ClinicalTrials.gov Identifier: NCT02588261

Date of Statistical Analysis Plan: Final Version 2.0, dated 11 Aug 2016

Sponsor: Astellas Pharma Global Development, Inc. (APGD)

1 Astellas Way

Northbrook, IL 60062

An Open-label, Randomized Phase 3 Efficacy Study of ASP8273 vs Erlotinib or Gefitinib in First-line Treatment of Patients with Stage IIIB/IV Non-small Cell Lung Cancer Tumors with EGFR Activating Mutations

ISN/Protocol 8273-CL-0302

ClinicalTrials.gov Identifier: NCT02588261

Date of Statistical Analysis Plan: Final Version 2.0, dated 11 Aug 2016

Sponsor: Astellas Pharma Global Development, Inc. (APGD)

1 Astellas Way

Northbrook, IL 60062

STATISTICAL ANALYSIS PLAN

Final Version 2.0, dated 11-August-2016

An Open-label, Randomized Phase 3 Efficacy Study of ASP8273 vs Erlotinib or Gefitinib in First-line Treatment of Patients with Stage IIIB/IV Non-small Cell Lung Cancer Tumors with EGFR Activating Mutations

ISN/Protocol: 8273-CL-0302

IND number: 119,902

Sponsor:

Astellas Pharma Global Development, Inc. (APGD)
1 Astellas Way
Northbrook, IL 60062

Table of Contents

List of In-Text Tables

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
AE	Adverse event
AGP	α-acid glycoprotein
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase (GPT)
ANC	Absolute neutrophil count
AntiHCV	Hepatitis C antibody
APEL	Astellas Pharma Europe Limited
APGD	Astellas Pharma Global Development
APTT	Activated partial thromboplastin time
AREC	Astellas Research Ethics Committee
AST	Aspartate aminotransferase (GOT)
AUC	Area under the curve
AUC_{inf}	Area under the concentration – time curve from time 0 extrapolated to infinity
AUC_{last}	Area under the concentration – time curve from time 0 up to the last quantifiable concentration
AUST	Astellas US Technologies
BUN	Blood urea nitrogen
CA	Competent authority
CIOMS	Council for International Organizations of Medical Sciences
CK	Creatine phosphokinase
CL/F	Apparent oral systemic clearance
CLtot	Total clearance
CNS	Central nervous system
CR	Complete response
CRF	Case report form
CRM	Continual reassessment method
CRO	Contract research organization
CSR	Clinical study report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
eCTD	Electronic Common Technical Document
CV	Coefficient of variation
CYP	Cytochrome P450
CYP3A4	Cytochrome P450 3A4
DCR	Disease control rate
DILI	Drug-induced liver injury
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EDC	Electronic data capture
EGFR	Epidermal growth factor receptor
eGFR	Estimated glomerular filtration rate
EORTC-QLQ-LC13	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Lung Cancer 13

Abbreviations	Description of abbreviations
EORTC-QLQ-C30	European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire
EQ-5D-5L	EuroQol 5-Dimension 5-Level Questionnaire
EU	European Union
FACT-EGFRI-18	Functional Assessment of Cancer Therapy – EGFRI 18 Questionnaire
FAS	Full analysis set
FFPE	Formalin Fixed Paraffin Embedded
GCP	Good Clinical Practice
GMP	Good Manufacturing Practice
GWAS	Genome-Wide Association Studies
HBsAg	Hepatitis B surface antigen
Hct	Hematocrit
HDL	High density lipoprotein
Hgb	Hemoglobin
HIPAA	Health Insurance Portability And Accountability Act
HIV	Human immunodeficiency virus
HNSTD	Highest nonseverely toxic dose
HR	Hazard ratio
IB	Investigator's Brochure
IC_{50}	50% inhibitory concentration
ICF	Informed consent form
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
ICR	Imprinting control region
IDAC	Independent Data Analysis Center
IDMC	Independent Data Monitoring Committee
IEC	Independent Ethics Committee
ILD	Interstitial lung disease
IMPD	Investigational Medicinal Product Dossier
IND	Investigational New Drug
INR	International normalized ratio
IRB	Institutional Review Board
IRR	Independent radiologic review
IRT	Interactive Response Technology
ISN	International Study Number
IUD	Intrauterine device
IUS	Intrauterine system
J-NDA	Japan New Drug Application
K_{inact}	Inactivation rate constant
KM	Kaplan-Meier
LA	Latin America
LA-CRF	Liver abnormality case report form
LDH	Lactate dehydrogenase
LDL	Low density lipoprotein
LFT	Liver function tests
LSO	Last subject out
MAA	Marketing Authorization Application
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration

Abbreviations	Description of abbreviations
MCV	Mean corpuscular volume
MDRD	Modification of Diet in Renal Disease
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
NA	North America
NASH	Nonalcoholic steatohepatitis
NCI	National Cancer Institute
NDA	New Drug Application
NOAEL	No observed adverse effect level
NSCLC	Non-small-cell lung cancer
ORR	Overall response rate
OS	Overall survival
PD	Progressive disease
PDAS	Pharmacodynamics analysis set
PET-CT	Positron emission tomography
PFS	Progression free survival
PFS#1	Disease progression on study drug
PFS#2	Progression free survival on next-line therapy
P-gp	Permeability-glycoprotein
PGx	Pharmacogenomic(s)
PHI	Protected health information
PKAS	Pharmacokinetic analysis set
PPS	Per protocol set
PR	Partial response
PRO	Patient-reported outcome
PT	Prothrombin time
QTcF	Fridericia corrected QT interval
QoL	Quality of life
RBC	Red blood cell
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
RP2D	Recommended phase 2 dose
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical Analysis Plan
SD	Stable disease
SOC	System organ class
SOP	Standard operating procedure
STD_{10}	Severely toxic dose in 10%
TBL	Total bilirubin
TEAE	Treatment-emergent adverse event
TKI	Tyrosine kinase inhibitor
TLF	Tables, listings and figures
t_{max}	The time after dosing when C_{max} occurs
TNM	TNM Classification of Malignant Tumors
TP	Total protein
ULN	Upper limit of normal
Vss	Steady state volume of distribution
Vz/F	Apparent volume of distribution

Abbreviations	Description of abbreviations
WBC	White blood cell
WHO	World Health Organization

List of Key Terms

Terms	Definition of terms
Baseline	Observed values/findings which are regarded observed starting point for comparison.
Enroll	To register or enter into a clinical trial. NOTE: Once a subject has been enrolled, the clinical trial protocol applies to the subject.
Intervention	The drug, therapy or process under investigation in a clinical study that is believed to have an effect on outcomes of interest in a study. (e.g., health-related quality of life, efficacy, safety, pharmacoeconomics).
Investigational period	Period of time where major interests of protocol objectives are observed, and where the test drug or comparative drug (sometimes without randomization) is usually given to a subject, and continues until the last assessment after completing administration of the test drug or comparative drug.
Post investigational period	Period of time after the last assessment of the protocol. Follow-up observations for sustained adverse events and/or survival are done in this period.
Screening period	Period of time before entering the investigational period, usually from the time of starting a subject signing consent until just before the test drug or comparative drug (sometimes without randomization) is given to a subject.
Randomization	The process of assigning trial subjects to treatment or control groups using an element of chance to determine assignments in order to reduce bias.
Screening	A process of active consideration of potential subjects for enrollment in a trial.
Screen failure	Potential subject who did not meet 1 or more criteria required for participation in a trial.
Study period	Period of time from the first site initiation date to the last site completing the study.
Variable	Any quantity that varies; any attribute, phenomenon or event that can have different qualitative or quantitative values.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary endpoints and other data.

The SAP is finalized and signed prior to any of the following: study unblinding, database hard lock, interim analysis, or accumulation of substantial amount of data in an open-label study to ensure lack of bias. For operational efficiency an earlier time is usually targeted and wherever possible, the SAP should be developed in parallel with protocol finalization. For Phase 2-4 studies the SAP should be developed and approved before First Subject In (FSI). If the expected interval between FSI and soft-lock is less than 12 weeks, then the SAP should be approved by 12 weeks prior to the planned date of soft-lock. If needed, revisions to the approved SAP may be made prior to database hard lock. Revisions will be version controlled.

This statistical analysis is coordinated by the responsible biostatistician of APGD-US. Any changes from the analyses planned in the SAP will be justified in the Clinical Study Report (CSR).

Prior to database hard lock, a final review of data and TLFs meeting will be held to allow a review of the clinical trial data and to verify the data that will be used for analysis set classification. If required, consequences for the statistical analysis will be discussed and documented. A meeting to determine analysis set classifications may also be held prior to database hard lock.

2 FLOW CHART AND VISIT SCHEDULE

Flow Chart:

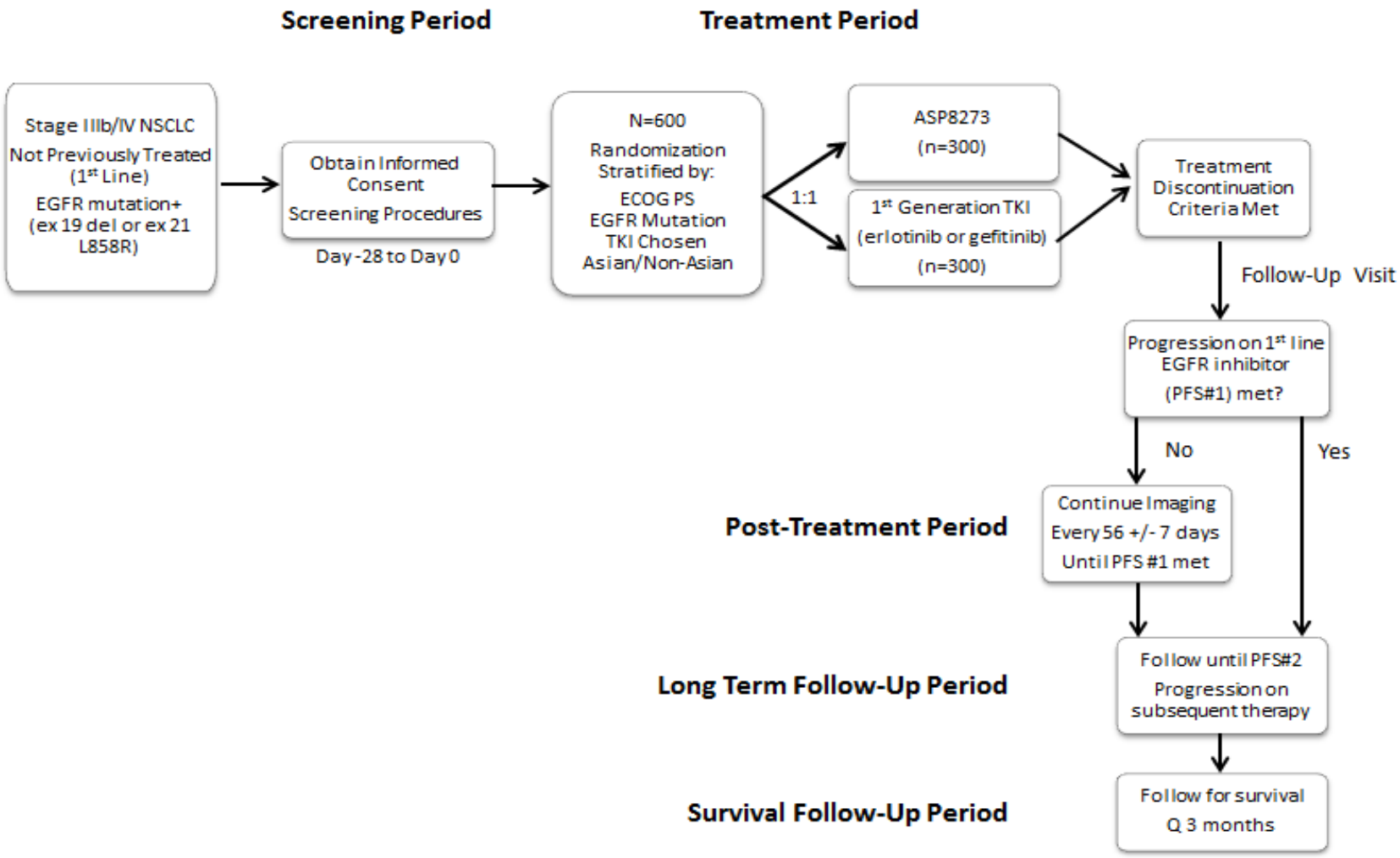

Table 1 Schedule of Assessments

VISIT / Base Date	Screening	Cycle 1		Cycle 2	Cycle 3-6	≥ Cycle 7 (odd cycles)	End of Treatment
		Day 1	Day 15	Day 1	Day 1	Day 1	Date of Last Dose
Visit Window	Day -28 to 0		±1 day	±3 days	±7 days	±7 days	+7 days
Informed Consent	X						
Medical and Disease History	X						
Confirmation of Eligibility	X	X					
Randomization[1]		X					
Study Drug Administration[2]		X	X	X	X	X	
Physical Examination [3]	X	X		X	X	X	X
ECOG PS	X	X		X	X	X	X
Vital Signs	X	X	X	X	X	X	X
Biochemistry[4]	X	X	X	X	X	X	X
Hematology[5]	X	X	X	X	X	X	X
Coagulation Parameters[6]		X		X	X	X	X
Urinalysis[7]		X	X	X	X	X	X
Serum Pregnancy Test[8]	X	X		X	X	X	X
12-lead ECG[9]	X	X	X	X	X	X	X
Image Assessment[10]	X				X	X	
Tumor Sample – EGFR[11]	X						
Post-progression Tumor Sample (optional)[12]							X
Plasma Sample – EGFR	X						X
Pharmacokinetics – Blood[13]			X	X	X	X	X
Plasma Sample –Biomarkers[14]		X			X	X	X
Whole Blood Sample – Biomarkers–		X					
Whole Blood Sample for PGx (optional)		X					
Concomitant Medication	X	X	X	X	X	X	X

Table continued on next page

VISIT	Screening	Cycle 1		Cycle 2	Cycle 3-6	≥ Cycle 7 (odd cycles)	End of Treatment
Base Date		Day 1	Day 15	Day 1	Day 1	Day 1	Date of Last Dose
Adverse Event	X	X	X	X	X	X	X
Quality of Life (QoL)[15]	X	X		X	X	X	X
Survival Assessment	X	X	X	X	X	X	X

[1] May be performed prior to C1D1 with prior sponsor approval

[2] Daily dosing, 28 days per cycle

[3] Physical examination and other evaluations include height (at screening only) weight, ECOG PS, and vital signs (pulse and blood pressure). Vital signs should be taken predose on dosing days. The physical exam only needs to be repeated on cycle 1 day 1 if clinically significant changes from Screening (in the opinion of the investigator) are observed.

[4] Biochemistry: See [Section 6.2, Table 3] for list of laboratory assessments. All biochemistry laboratory tests should be conducted after the subject has been fasting for at least 6 hours in order to ensure accurate interpretation of glucose values. Fasting status should be recorded in source documents. Biochemistry tests will be sent to a central laboratory for analysis. Additional assessments may be done centrally or locally to monitor adverse events or as required by dose modification requirements.

[5] Hematology: see [Section 6.2, Table 3] for list of laboratory assessments. Hematology tests will be sent to a central laboratory for analysis,

[6] Coagulation: see [Section 6.2, Table 3] for list of laboratory assessments. Coagulation tests will be taken on day 1 of cycles 1-3 and every odd cycle thereafter (cycles 5, 7, 9 etc.). Coagulation tests will be sent to a central laboratory for analysis.

[7] Urinalysis: see [Section 6.2, Table 3] for list of urinalysis assessments. Urinalysis tests will be sent to a central laboratory for analysis.

[8] For female subjects of child bearing potential only. Serum pregnancy tests every cycle where a visit takes place will be done throughout the study.

[9] ECG time points are: Screening; cycle 1 day 1 predose; cycle 1 day 15 predose; and day 1 predose of cycle 2 and each subsequent cycle where a visit takes place. When collected on the same day, ECG should be collected prior to pharmacokinetic samples.

[10] Image assessments will be done every 56 days (± 7 days), scheduled in a way to allow results to be available for the odd cycle day 1 visit (i.e., prior to cycles 3, 5, 7, 9, etc.). Imaging of the brain is required at baseline for all subjects, and may be repeated throughout the study as clinically indicated. If a brain lesion is selected as a target lesion at baseline, the assessment should be repeated at all subsequent imaging timepoints. CT scan with contrast (chest and abdomen) is the preferred modality for tumor assessment. MRI is acceptable if local standard practice or if CT scans are contraindicated in a subject (e.g., subject is allergic to contrast media). All other RECIST-approved scanning methods such as X-ray are optional. PET-CT scans must use contrast. To ensure comparability, the screening and subsequent assessment of response should be performed using identical techniques. The same method should be employed and assessed by the same individual on each occasion if possible. Imaging assessments, including brain, should utilize the same methods throughout the study. Confirmation scans for CR or PR should be done at the next scheduled assessment. A chest x-ray or other appropriate imaging of the lungs should be performed in addition to specified imaging time points if a subject develops symptoms suggestive of ILD.

Footnotes continued on next page

[11] Eligibility may be determined by a previously documented EGFR mutation test result. However, there should be sufficient tumor tissue from the specimen used to generate the EGFR test result to send to the central lab for confirmatory testing. If a subject does not have a previously documented EGFR mutation test result, then a fresh specimen must be obtained and sent to the central lab for eligibility testing. If slides are submitted, the slides should be freshly cut from the FFPE block within 18 days of sending to the central lab. A plasma sample for EGFR mutation detection will also be collected and submitted to the central laboratory.

[12] For subjects who signed a separate ICF, an optional post-progression tumor sample for exploratory biomarker analysis should be collected following local or central confirmation of disease progression and prior to commencement of subsequent anticancer therapy.

[13] Pharmacokinetic samples of ASP8273 will be taken for subjects randomized to Arm A (ASP8273) at predose on cycle 1 day 15 and predose on day 1 of cycles 2, 3, 5, 9 and 13 and at the End of Treatment visit. The date and time of each blood sample collection will be recorded to the nearest minute. In addition, the date and time of the ASP8273 doses taken on the pharmacokinetic days and the last 2 ASP8273 doses taken prior to each pharmacokinetic day will be recorded to the nearest minute. In the event ASP8273 treatment has been withheld for ≥3 days prior to the timepoint, the PK sample will not be collected.

[14] Plasma samples for biomarkers will be collected at predose on day 1 of each odd cycle (cycles 1, 3, 5, 7, 9 etc.) and at the End of Treatment visit. Whole blood sample for biomarkers will be collected once at baseline.

[15] QoL questionnaires are to be administered on visit days before any other scheduled assessments are conducted and before the disease status is discussed with the subject.

Table 2 Post-treatment Discontinuation Schedule of Assessments

VISIT	Follow-up Visit	Post-treatment Follow-up Period	Long-term Follow-up Period	Survival Follow-up Period
Base Date	**Date of last Dose +30 days**	**Every 56 days**	**Every 3 Months**	**Every 3 months**
Visit Window	+7 days	±7 days	±7 days	±7 days
ECOG PS	X			
Vital Signs	X			
Image Assessment		X[1]		
Subsequent therapy assessment[2]	X	X	X	
QoL	X			
QoL – EQ-5D-5L only[3]		X	X	
Survival Assessment	X	X	X	X[4]
Adverse Event	X	X		

[1] If a subject discontinues study drug prior to radiographic disease progression (i.e., PFS#1), the subject should continue to undergo imaging assessment (including brain imaging when indicated and collection of tumor measurements) every 56 days (±7 days) in the posttreatment follow-up period until PFS#1 is documented per the investigator, the subject misses 2 or more consecutive scheduled tumor assessments, or the subject starts another cancer treatment, whichever occurs earlier.

[2] After PFS#1, subjects will be followed in the long-term follow-up period per institutional guidelines, but not less frequently than every 3 months to confirm survival status and collect subsequent anticancer treatment details and progression status until PFS#2 is documented or the subject starts another cancer treatment, whichever occurs earlier. Phone contact with subject is sufficient for follow-up. Additional follow-up contacts may be required per sponsor request for analysis purposes.

[3] Quality of life (EQ-5D-5L only) will be assessed at each of the posttreatment or long-term follow-up survival contacts for 6 months after treatment discontinuation. QoL may be assessed by telephone module if follow-up contact is done by phone.

[4] Subjects will be followed via phone call in the survival follow-up period approximately every 3 months to collect survival status until subject death or study closure. Additional follow-up contacts may be required per sponsor request for analysis purposes.

3 STUDY OBJECTIVE(S) AND DESIGN

3.1 Study Objective(s)

3.1.1 Primary Objective

- To evaluate the progression free survival (PFS), based on independent radiologic review (IRR), of ASP8273 compared to erlotinib or gefitinib in patients with locally advanced, metastatic or unresectable stage IIIB/IV adenocarcinoma non-small cell lung cancer (NSCLC) with EGFR activating mutations

3.1.2 Secondary Objectives

- Overall survival (OS)
- Overall response rate (ORR) as assessed by IRR
- PFS as assessed by the investigator
- Disease control rate (DCR) as assessed by IRR
- Safety of ASP8273
- To evaluate Quality of Life (QoL) and patient-reported outcome (PRO) parameters as measured by FACT-EGFRI-18, EORTC-QLQ-LC13, EORTC-QLQ-C30 and EQ-5D-5L

3.1.3 Exploratory Objectives

- To evaluate potential biomarkers that may affect treatment outcome
- Evaluation of the pharmacokinetics of ASP8273
- To evaluate the impact of subsequent therapy on progression free survival on next-line therapy (PFS#2)

3.2 Study Design

3.2.1 Study Design

This multinational, open-label randomized study will evaluate ASP8273 compared to erlotinib or gefitinib (1st generation EGFR TKI) as first line therapy in subjects with locally advanced, metastatic or unresectable stage IIIB/IV adenocarcinoma NSCLC (newly diagnosed or recurrent) with EGFR activating mutations (exon 19 deletion or exon 21 L858R), with or without a T790M mutation, who have not previously been treated with an EGFR inhibitor (1st line).

PFS by IRR is the primary variable. Secondary variables include OS, ORR, PFS by the investigator, DCR, safety and QoL/PRO. Exploratory variables include biomarkers, pharmacokinetics and PFS#2 on subsequent therapy.

Approximately 600 subjects will be randomized 1:1 to 1 of 2 treatment arms (Arm A or Arm B). Arm A will receive 300 mg daily of ASP8273 and Arm B will receive either 150 mg erlotinib or 250 mg gefitinib daily, as decided by the investigator prior to randomization. Both arms will follow 28-day cycles of continuous dosing. Subjects will be stratified according to the following: ECOG PS (0, 1 or 2), EGFR mutation status (exon 19

deletion or mutations in exon 21 [L858R]), TKI chosen (erlotinib or gefitinib) and race (Asian versus non-Asian).

Screening will take place up to 28 days prior to subject randomization on cycle 1 day 1. Subjects will start with cycle 1 and continue on to subsequent 28-day cycles until 1 of the discontinuation criteria are met. Subjects will visit the clinic for evaluations on day 1 and day 15 of treatment cycle 1 and then on day 1 of subsequent treatment cycles up to cycle 7. Subjects who continue past cycle 7 will have visits on day 1 of each odd-numbered cycle.

Quality of life will be assessed during each treatment cycle where a visit takes place, at the follow-up visit, and at each of the post-treatment or long-term follow-up survival contacts for 6 months after treatment discontinuation (EQ-5D-5L only).

Imaging will be evaluated at baseline and every 56 days (± 7 days) throughout the study. A blinded independent central reader will review all imaging scans to confirm complete response (CR), partial response (PR), stable disease (SD) or progressive disease (PD) per Response Evaluation Criteria in Solid Tumors (RECIST) V1.1. Disease progression on study drug (PFS#1) should be confirmed by the blinded independent central reader before subjects are discontinued from treatment.

Following discontinuation from study drug, subjects will have a follow-up visit within 30 days after their last dose of drug for safety assessments. If a subject discontinues study drug prior to radiographic disease progression (i.e., PFS#1), the subject should enter the post-treatment follow-up period and continue to undergo imaging assessments every 56 days (+/– 7 days) until PFS#1 is documented or the subject starts another anticancer treatment, whichever occurs earlier.

Following PFS#1, subjects will enter the long-term follow-up period and be followed every 3 months from the date of the follow-up visit for survival status and progression status on subsequent therapy (i.e., PFS#2). All subsequent anticancer therapy including date and site of progression for PFS#2 will be recorded on the case report form. Subjects will be followed until PFS#2 is documented such as investigator-determined progression, death, subject has ended the subsequent anticancer therapy or the subject starts another anticancer treatment, whichever occurs earlier.

Following PFS#2, subjects will enter the survival follow-up period and be followed every 3 months for survival status.

Samples for pharmacokinetics, biomarkers and pharmacogenomics (PGx) as well as formalin fixed paraffin embedded (FFPE) specimens for central analysis will be collected. Eligibility may be determined by a previously documented EGFR mutation result. However, there should be sufficient tissue from the specimen used to generate the EGFR test result to send to the central lab for confirmatory testing. If a subject does not have a previously documented EGFR mutation test result, then a fresh specimen must be obtained and sent to the central lab for eligibility testing. From initial progression (PFS#1), an optional post-progression tumor tissue sample for exploratory biomarker analysis may be collected following centrally confirmed disease progression for subjects who sign a separate informed consent form (ICF).

An independent Data Monitoring Committee (IDMC) will be chartered to oversee safety and the planned futility analysis which will occur after at least 210 PFS events have been observed.

The primary analysis will occur when at least 420 PFS#1 events are observed which is expected to occur approximately 36 months after randomization. Primary endpoint, secondary endpoints and other endpoints will be analyzed at the time of primary analysis.

3.2.2 ASP8273 Dose/Dose Regimen and Administration Period

ASP8273 300 mg and erlotinib 150 mg will be taken orally once daily, on an empty stomach defined as no food for at least 2 hours before and 1 hour after dosing, at approximately the same time(s) every day.

Gefitinib 250 mg will be taken orally once daily, with or without food, at approximately the same time every day.

Concomitant medications should not be administered within 2 hours before or after dosing with ASP8273, erlotinib or gefitinib.

If a study drug dose is missed, it should be taken as soon as the subject remembers unless more than 12 hours has passed since the intended dosing time, in which case the subject should wait and take the next scheduled dose. Subjects should not make up a missed dose by taking a double dose at the next scheduled dose.

3.3 Randomization

Randomization in a 1:1 ratio to each of the 2 treatment arms will be performed via Interactive Response Technology (IRT). Randomization will be stratified by ECOG PS (0, 1 or 2), EGFR mutation type (exon 19 deletion or mutation in exon 21 L858R), TKI chosen (erlotinib versus gefitinib) and race (Asian versus non-Asian). During screening for each subject the site will need to designate in IRT if the subject will receive erlotinib or gefitinib in the event they are randomized to the comparator arm. Prior to the initiation of the study treatment, the site staff will contact the IRT in order to determine the randomly assigned treatment arm. Specific procedures for randomization through the IRT are contained in the study procedures manual.

4 SAMPLE SIZE

Approximately 600 subjects will be randomized in a 1:1 ratio to 2 treatment arms: Arm A (ASP8273) and Arm B (erlotinib or gefitinib). Randomization will be stratified by ECOG PS (0, 1 or 2), EGFR mutation type (exon 19 deletion or mutation in exon 21 [L858R]), TKI chosen (erlotinib versus gefitinib) and race (Asian versus non-Asian). Assuming HR=0.667 (median PFS in Arm A and Arm B are 15.6 months and 10.4 months, respectively), a total of 420 PFS events will provide approximately 95% power to detect a statistically significant difference at type I error rate of 1-sided 0.025.

5 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

Detailed criteria for analysis sets will be laid out in Classification Specifications (CS) and the allocation of subjects to analysis sets will be determined prior to database hard lock.

Full Analysis Set (FAS) and Per Protocol Set (PPS) will be used for efficacy analysis. Safety Analysis Set (SAF) will be used for the analyses of safety variables. Pharmacokinetic Analysis Set (PKAS) will be used for pharmacokinetic analyses. The biomarker pharmacodynamics analysis set (PDAS) will be used for all analyses of pharmacodynamics data.

The data from all patients who were enrolled in the study will be included in the data listings. All enrolled subjects are those who signed the informed consent form and were assigned a subject number.

5.1 Full Analysis Set (FAS)

The full analysis set will consist of all subjects who are randomized. This will be the primary analysis set for efficacy analyses.

5.2 Per Protocol Set (PPS)

The per protocol set will consist of the subset of the FAS who do not meet criteria for PPS exclusion. These criteria are to capture relevant non-adherence to the protocol. In addition, subjects in the PPS are required to have both baseline imaging and at least 1 post-baseline imaging scan. The PPS will be a secondary analysis set for efficacy analyses. Select demographic and baseline characteristics may also be summarized for the PPS.

5.2.1 Reasons for Exclusion From PPS

The following reasons may lead to subject's exclusion from PPS:

- Entered into the study even though the subject violates the inclusion or exclusion criteria which may affect the assessment of the efficacy of the study drug
    - Inclusion criteria that affect efficacy are: #3, #9, #10, #12, #13 and #14.
    - Exclusion criteria that affect efficacy are: #1, #2, #3, #5, #9, #12, #18, #19 and #21.

  For inclusion/exclusion criteria details, refer to Section 3.2 and 3.3 of the study protocol.
- Received study treatment dose/s other than permitted by protocol for more than 14 days
- Administration of anti-NSCLC treatment prohibited by protocol
- No baseline or post-baseline imaging assessment

5.3 Safety Analysis Set (SAF)

For the statistical summary of the safety data, the safety analysis set (SAF) will be used. The SAF consists of all subjects who take at least 1 dose of study medication, and will be used for safety analyses.

5.4 Pharmacokinetics Analysis Set (PKAS)

The pharmacokinetic analysis set (PKAS) consists of the subset of the SAF population for which at least one plasma concentration data is available for whom the time of dosing on the previous day of sampling is known. Additional subjects may be excluded from the PKAS at the discretion of the pharmacokineticist. Any formal definitions for exclusion of subjects or time-points from the PKAS will be documented in the Classification Specifications and determined in the Classification Meeting.

5.5 Pharmacodynamics Analysis Set (PDAS)

The biomarker pharmacodynamics analysis set (PDAS) will include the subjects from the SAF population for whom sufficient pharmacodynamic measurements were collected. The PDAS will be used for all analyses of pharmacodynamics data.

6 ANALYSIS VARIABLES

6.1 Efficacy Endpoints

6.1.1 Primary Efficacy Endpoint

The primary endpoint is PFS assessed by the blinded IRR. For each subject, PFS is defined as the time from the date of randomization until the date of radiological disease progression (i.e., PFS#1) assessed by IRR, or until death due to any cause. If a subject has neither progressed nor died, the subject will be censored at the date of last radiological assessment or at the date of randomization if no post-baseline radiological assessment is available. Subjects who receive any further anticancer therapy for the disease before radiological progression will be censored at the date of the last radiological assessment before the anticancer therapy started. If progression or death occurs after missing 2 scheduled radiological assessments, the subject will be censored at the date of last radiological assessment or at the date of randomization if no post-baseline radiological assessment is available. Missing 2 scheduled imaging assessments is defined as the gap between two imaging assessments is more than 140 days (2.5 times the scheduled imaging assessment interval (2.5 x 56 days)).

PFS (in days) is calculated as:

(Date of death or documented Progression or censoring) – (Date of randomization) +1.

To apply the cut-off date to PFS is to exclude tumor assessments, death and anti-cancer therapy date after cut-off date in the analysis.

Note: Patient cannot be censored at “Not Evaluable (NE)”. If NE is the only previous assessment, then PFS will be censored at Day 1.

6.1.2 Secondary Efficacy Endpoints

6.1.2.1 Overall Survival (OS)

OS is defined as the time from the date of randomization until the documented date of death from any cause. All events of death will be included, regardless of whether the event occurred while the subject is still taking study drug or after the subject discontinue study drug. Subjects who are still alive at the time of analysis will be censored at the last day known to be alive or at the analysis cutoff data, whichever is earlier. Subjects died after the analysis cut-off date will be censored at the analysis cut-off date.

OS (in days) is calculated as:

(Date of death or known to be alive) – (Date of randomization) +1.

6.1.2.2 Objective Response Rate (ORR)

Best overall response is determined once all tumor response data for the subject is available. Subjects will be classified by best response on study as outlined in RECIST v1.1 criteria. For best overall response of stable disease (SD), SD must be documented as present at least once after study entry and at least 7 weeks (the minimum required days for first scheduled imaging assessment after first dose) after first dose.

ORR is defined as the proportion of subjects with best overall response as complete response (CR) or partial response (PR) without confirmation based on the RECIST v1.1 as assessed by the blinded IRR. ORR as assessed by the local investigator will be performed as a sensitivity analysis.

Analysis of ORR with confirmation defined as the proportion of subjects with best overall response as confirmed CR or confirmed PR based on the RECIST v1.1 as assessed by the blinded IRR may also be performed. Confirmation of CR or PR should occur at the next scheduled assessment (not less than 4 weeks following the initial assessment at which CR or PR is observed)

6.1.2.3 Disease Control Rate (DCR)

The DCR is defined as the proportion of subjects with a CR or PR (without confirmation) or SD based on RECIST v1.1 as assessed by the blinded IRR. DCR as assessed by the local investigator will be performed as a sensitivity analysis.

Analysis of DCR with confirmation defined as the proportion of subjects with best overall response as confirmed CR, confirmed PR or SD based on the RECIST v1.1 as assessed by the blinded IRR may also be performed.

6.1.2.4 Duration of Response (DOR)

DOR is defined as the time from the date of the first response CR/PR (whichever is first recorded) as assessed by IRR to the date of radiographical progression or date of censoring. If a subject has not progressed, the subject will be censored at the date of last radiological assessment or at the date of first CR/PR if no post-baseline radiological assessment is

available. DOR will be derived for subjects with best overall response as CR or PR (without confirmation).

DOR (in days) is calculated as:

(Date of documented progression or censoring) – (Date of the first response CR/PR) +1.

To apply the cut-off date to DOR is to exclude tumor assessments after cut-off date in the analysis.

6.1.2.5 PFS#1 as assessed by Local Investigator

See Section 6.1.1 for the definition of PFS#1. Tumor assessments as assessed by local investigator will be used in the derivation of PFS.

6.1.3 Exploratory Efficacy Endpoint

6.1.3.1 PFS on Next-line therapy (PFS#2)

PFS#2 is defined as the time from initiation of new systemic anti-cancer treatment (excluding local radiation, surgical removal of metastatic site or protocol defined Erlotinib or Gefitinib or ASP8273 based on initial treatment allocation per subject) to investigator-determined progression , discontinuation of new treatment, start of a different treatment or death from any cause, whichever comes first. Otherwise, subjects will be censored.

Subjects who have not started new systemic anti-cancer treatment will be censored at Day 1. Subjects who are alive, started new systemic anti-cancer treatment and for whom a PFS#2 event has not been observed should be censored at the last time known to be alive.

PFS#2 (in days) is calculated as:

(Date of death or investigator-determined progression on next-line therapy or discontinuation of next-line therapy or started another therapy or censoring) – (Date of randomization) +1.

6.2 Safety Variables

Safety endpoints such as AEs, laboratory tests, vital signs, ECGs and ECOG performance status are secondary endpoints of the study.

Safety will be assessed by evaluation of the following variables:

- Treatment-emergent adverse events (TEAEs; frequency, severity, seriousness, and relationship to study drug).
    - TEAE is defined as an adverse event observed after starting administration of the study drug. If the adverse event occurs on Day 1 and the onset check box is marked “Onset after first dose of study drug” or the onset check box is left blank, then the adverse event will be considered treatment emergent. If the adverse event occurs on Day 1 and the onset check box is marked “Onset before first dose of study drug”, then the adverse event will not be considered treatment emergent. If a subject experiences an event both during the pre-

investigational period and during the investigational period, the event will be considered as TEAE only if it has worsened in severity (i.e. it is reported with a new start date). All adverse events collected that begin within 30 days after taking the last dose of study drug will also be counted as TEAE. Any AEs with onset dates completely missing will be considered TEAEs in summaries. AEs with partially missing onset dates will be assumed TEAEs unless the available portion of the date indicates that the onset was strictly before start of study medication.

- A drug-related TEAE is defined as any TEAE with at least possible relationship (possibly or probably related) to study treatment as assessed by the investigator or with missing assessment of the causal relationship

- Serious adverse events (SAEs) include adverse events that are flagged as serious by the investigator on eCRF, or upgraded by the Sponsor based on review of the Sponsor's list of Always Serious term.

- Clinical laboratory variables

Below is a table of the laboratory tests that will be performed during the conduct of the study. Refer to the Schedule of Assessments for study visit collection dates.

Additional laboratory tests should be performed according to institutional standard of care. Clinical significance of out-of-range laboratory findings is to be determined and documented by the investigator or delegated sub-investigator who is a qualified physician.

Table 3 Laboratory Assessments

Panel/Assessment	**Parameters to be analyzed**	**NCI-CTCAE Grading (Hyper/Hypo)**
Hematology	Hematocrit (Hct)	N/A
	Hemoglobin (Hgb)	Both
	Red Blood Cell Count (RBC)	N/A
	White Blood Cell Count (WBC)	Both
	WBC differential	N/A
	Platelets	Hypo
	Mean Corpuscular Volume (MCV)	N/A
	Mean Corpuscular Hemoglobin (MCH)	N/A
	Mean Corpuscular Hemoglobin Concentration (MCHC)	N/A
Table continued on next page		

Panel/Assessment	Parameters to be analyzed	NCI-CTCAE Grading (Hyper/Hypo)
Biochemistry (fasting)	Sodium (Na)	Both
	Magnesium (Mg)	Both
	Creatine Phosphokinase (CK)	Hyper
	Potassium (K)	Both
	Calcium (Ca)	Both
	Chloride (Cl)	N/A
	Phosphate (P)	Hypo
	Creatinine (Cr)	Hyper
	Glucose (Gl)	Both
	Blood Urea Nitrogen (BUN)	N/A
	Alkaline Phosphatase (ALP)	Hyper
	Aspartate Aminotransferase (AST)	Hyper
	Alanine Aminotransferase (ALT)	Hyper
	Lactate Dehydrogenase (LDH)	N/A
	Bilirubin Total (TBL) (total and direct)	Hyper
	Total Protein (TP)	N/A
	Albumin (Alb)	Hypo
	Bicarbonate (HCO3)	N/A
	Serum HCG for female subjects of childbearing potential	N/A
Urinalysis	Protein	N/A
	Glucose	N/A
	pH	N/A
	Occult Blood	N/A
	Ketones	N/A
	Bilirubin	N/A
	Urobilinogen	N/A
	Sodium	N/A
	Potassium	N/A
	Chloride	N/A
Coagulation Profile	PT (sec)	N/A
	APTT	Hyper
	INR	Hyper

- Vital signs (systolic and diastolic blood pressures (mmHg), radial pulse rate (beats/minute) and body temperature)
- 12-lead electrocardiogram (ECG)
- ECOG performance scores
- Physical Examination Standard(general appearance, skin, eyes, ears, nose, throat, neck, cardiovascular, chest and lungs, abdomen, musculoskeletal, neurologic status, mental status, and lymphatic systems)

6.3 Pharmacokinetics Variables

Plasma concentration data of ASP8273 and its metabolite(s) (if applicable) will be analyzed and descriptive statistics will be provided. Additional model-based analyses may be performed and reported separately.

6.4 Biomarker Variables

Tumor tissue sample (FFPE) will be obtained and sent for central confirmatory testing to evaluate for EGFR activating mutation. The sample sent for confirmatory testing should be from the same specimen as used for local testing for eligibility. A plasma sample will be collected to evaluate EGFR mutations. Plasma and whole blood samples will be collected for exploratory analysis of biomarkers. An optional post-progression tumor tissue sample for exploratory biomarker analysis may be collected following centrally confirmed disease progression for subjects who sign a separate ICF.

6.5 Quality of Life Variables

Subjects will be asked to complete the QoL and PRO questionnaires electronically at baseline and at day 1 of each cycle where a visit takes place.

6.5.1 FACT-EGFRI-18

The Functional Assessment of Chronic Illness Therapy-EGFR Inhibitors Subscale (FACT-EGFRI-18) is a treatment-specific subscale of the Functional Assessment of Chronic Illness Therapy (FACIT) measurement system specifically designed to assess the effect of EGFR inhibitors on quality of life. (Wagner et al). The FACT-EGFRI-18 is an 18-item Likert-scaled questionnaire, arranged in three HRQL dimensions: physical (seven items), social/emotional (six items), and functional well-being (five items). The instrument takes less than 10 minutes to complete, and the recall period is the past 7 days. The response scores ranged from 0 to 4, and the response categories include “not at all”, “a little bit”, “somewhat”, “quite a bit”, and “very much”. Negatively worded items (e.g., “My skin bleeds easily “or “My skin condition affects my mood”) are reverse-scored, so that participants who experience a higher impact of symptom burden on HRQL receive a lower score (range 0-72).

6.5.2 EORTC-QLQ-C30

The EORTC-QLQ-C30 is a 30-item cancer-specific instrument (Aaronson et al 1993). Multi-trait scaling was used to create five functional domain scales: Physical, Role, Emotional, Social and Cognitive; two items evaluate global QoL; in addition, three symptom scales assess Fatigue, Pain and Emesis; and six single items assess other symptoms.

Validity for EORTC-QLQ-C30 was illustrated in a study evaluating quality of life in a sample of NSCLC patients, where the EORTC-QLQ-C30 discriminated between participants in the intervention arm (186 patients undergoing adjuvant chemotherapy) and in the observation arm (173 patients), with significant differences in the domains of Global quality of life, Physical functioning, Role functioning and Social functioning showing better scores for the observation arm at 3 months, returning to baseline score at 9 months (Bezjak et al 2008). In the most recent version of the instrument (version 3), the Cronbach's alpha-

coefficients for internal consistency ranged from 0.56 to 0.85, with emotional functioning having the highest Cronbach's alpha-coefficient (Cankurtaran et al 2008). The EORTC-QLQ-C30 discriminated between different types of cancers (lung, colorectal and prostate) in a study exploring peri-diagnostic and survival wait times (Grunfeld et al 2009). Predictive validity of the EORTC-QLQ-C30 was illustrated in a study with 538 cancer patients (101 had lung cancer) where Function and Symptom scales were predictive of overall patient satisfaction upon univariate analysis (Lis et al 2009).

The instrument's evidence for the ability to detect change has been reported for NSCLC treatment trials (Yang et al 2013 & O'Brien et al 2004). Evidence supporting interpretation of scores (i.e., availability of MID estimates in patients with NSCLC) also is available (Maringwa et al 2011). Like the other EORTC instruments, the recall period for items is the past 7 days, and response options include a 4-point Likert scales, yes/no options, or a 7-point scale for 2 global questions related to overall health status and HRQL. The instrument takes less than ten minutes to complete. The total score for the instrument ranges from 0 to 100, with a high score for a functional scale representing a high/healthy level of functioning and a high score for a symptom scale or item representing a high level of symptomatology or problems (Fayers et al 2001).

6.5.3 EORTC-QLQ-LC13

The EORTC-QLQ-LC13 is a validated module of the EORTC-QLQ-Core 30, developed in 1994. Module items evaluate symptoms such as cough, hemoptysis, shortness of breath, sore mouth or tongue, dysphagia, tingling hands or feet, hair loss and pain. The developers of EORTC-QLQ-LC13 indicate that it is highly preferred not to use the LC13 alone (without the core module QLQ-C30), since the module has been designed to be used together with the core questionnaire, and the content validity is based upon this combination. The response options and scoring system are the same as for the EORTC QLQ-C30, and the administration is similar. The recall period for items is the past 7 days, and response options include either a 4-point Likert scale or yes/no options. The instrument takes less than ten minutes to complete. The total score for the instrument ranges from 0 to 100. A high score for a functional scale represents a high/healthy level of functioning whereas a high score for a symptom scale or item represents a high level of symptomatology or problems (Fayers et al 2001).

6.5.4 EQ-5D-5L

The EQ-5D is a generic preference-based measure that indirectly measures the utility for health that generates an index-based summary score based upon societal preference weights (Pickard et al 2007). The EQ-5D-5L consists of 6 items that cover 5 main domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a general visual analog scale for health status. The utility scores allow for comparisons of disease burden to be made across various conditions, and the calculated quality-adjusted life years as an outcome for determining the cost effectiveness of health care interventions. The index-based score is typically interpreted along a continuum where 1 represents best possible health and 0 represents dead.

6.6 Other Variables

- Body Mass Index (BMI)

  BMI = weight (kg) / [height (m)]^2

- Duration of exposure

  Duration of exposure to a study drug will be calculated in days, using the following formula:

  (Date of last dose of study drug – Date of first dose) + 1

  When the start or stop date is missing, then the exposure will be treated as missing.

- Cumulative Actual Dose

  Total amount of study drug actually taken by the patient from first dose date to last dose date

- Planned Dose Intensity

  Planned dose intensity for ASP8273 is 300 mg/day, for erlotinib is 150 mg/day and for gefitinib is 250 mg/day.

- Actual Dose Intensity

  Defined as the cumulative actual dose divided by duration of exposure

- Relative Dose Intensity (RDI)

  $$\frac{\text{Actual Dose Intensity}}{\text{Planned Dose Intensity}} \times 100$$

- Duration of NSCLC

  Duration of NSCLC will be calculated in days using the following formula:

  (Randomization date – date of initial diagnosis of NSCLC) + 1

  If the date of initial diagnosis of NSCLC is incomplete or missing, then the duration of NSCLC will be missing.

- Duration of NSCLC since date of locally advanced or metastatic disease diagnosis

  Duration of NSCLC since date of locally advanced or metastatic disease diagnosis will be calculated in days using the following formula:

  (Randomization date – date of locally advanced or metastatic disease diagnosis) + 1

If the date of locally advanced or metastatic disease diagnosis is incomplete or missing, then the duration of NSCLC since date of locally advanced or metastatic disease diagnosis will be missing.

- Previous and concomitant medication

  Previous medication is defined as medication with at least one dose taken before the date of the first dose of study drug.

  Concomitant medication is defined as medication with at least one dose taken between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug.

- Previous and concomitant non-medication therapy

  Previous non-medication therapy is defined as non-medication therapy administered at least once before the date of the first dose of study drug.

  Concomitant medication is defined as non-medication therapy administered at least once between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug.

7 STATISTICAL METHODOLOGY

7.1 General Considerations

For continuous variables, descriptive statistics will include the number of subjects (n), mean, standard deviation, median, minimum and maximum. When needed, the use of other percentiles (e.g., 10%, 25%, 75% and 90%) will be mentioned in the relevant section. Frequencies and percentages will be displayed for categorical data. Percentages by categories will be based on the number of subjects with no missing data, i.e. will add up to 100%. Kaplan-Meier survival curves will be displayed for time-to-event variables and median survival time will be estimated with 2-sided 95% confidence interval (CI).

Summaries based on FAS and PPS (e.g. disposition, baseline and efficacy data) will be presented by planned treatment, unless specifically stated otherwise. Safety analysis and other summaries based on SAF, PKAS or PDAS will be presented by actual treatment received.

All statistical comparisons will be made using one sided test at the α=0.025 significance level unless specifically stated otherwise. All null hypotheses will be Arm A is not better than Arm B, all alternative hypotheses will be Arm A is better than Arm B, unless specifically stated otherwise.

All data processing, summarization, and analyses will be performed using SAS® Version 9.2 or higher on Unix. Specifications for table, figure, and data listing formats can be found in the TLF specifications for this study.

Study day for safety assessments (e.g. laboratory assessment, onset of adverse events, vital signs, etc.) will be calculated in reference to the first dose date. For assessments conducted before the first dose, study day will be calculated as (assessment date – first dose date). For assessments conducted on or after the first dose, study day will be calculated as (assessment date – first dose date + 1).

Study day for efficacy assessments (imaging assessment, overall survival, tumor response, etc.) will be calculated in reference to the randomization date. For assessments conducted before randomization, study day will be calculated as (assessment date – randomization date). For assessments conducted on or after randomization, study day will be calculated as (assessment date – randomization date + 1).

For efficacy evaluation, baseline is defined as the last available measurement before randomization. For safety evaluation, baseline is defined as the last available measurement before first dose. Unless otherwise specified, all summaries will be presented by treatment arm.

For the definition of subgroups of interest please refer to Section 7.8

7.2 Study Population

7.2.1 Disposition of Subjects

The following subject data will be presented:

- Number and percentage of subjects with informed consent, discontinued before randomization, randomized (overall only);
- Number and percentage of subjects randomized in each analysis set, by treatment group and overall;
- Number and percentage of subjects discontinued treatment, by primary reason for treatment discontinuation, by treatment group for FAS and SAF;
- Number and percentage of subjects completed the 30 day follow-up visit, by primary reason for 30 day follow-up discontinuation for randomized subjects, by treatment group;
- Number and percentage of subjects completed or discontinued post-treatment period, by primary reason for post-treatment discontinuation for randomized subjects, by treatment group;
- Number and percentage of subjects completed or discontinued long term follow-up period, by primary reason for long term follow-up discontinuation for randomized subjects, by treatment group;
- Number and percentage of subjects completed or discontinued survival follow-up period, by primary reason for survival follow-up discontinuation for randomized subjects, by treatment group;
- Number and percentage of screen failure subjects, by primary reason for screen failure for screen failure subjects only;
- Number and percentage of subjects excluded from PPS by reason for exclusion defined in Section 5.2.1, by treatment arm for FAS.

7.2.2 Protocol Deviations

Protocol deviations as defined in the study protocol (Section 8.1.6 Protocol Deviations) will be assessed for all subjects randomized. The number and percentage of subjects meeting any criteria will be summarized for each criterion and overall, by treatment group and overall as well as by study site. Subjects deviating from a criterion more than once will be counted once for the corresponding criterion. Any subjects who have more than one protocol deviation will be counted once in the overall summary. A data listing will be provided by site and subject.

The protocol deviation criteria will be uniquely identified in the summary table and listing. The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria,

PD2 - Developed withdrawal criteria during the study and was not withdrawn,

PD3 - Received wrong treatment or incorrect dose,

PD4 - Received excluded concomitant treatment.

7.2.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized by descriptive statistics.

Number and percentage of subjects randomized in each country and site will be presented by treatment group and overall for the FAS.

Descriptive statistics for age, weight, body mass index (BMI) and height at study entry will be presented. Frequency tabulations for sex, ethnicity, age group (defined in Section 7.8) and race will be presented. Stratification factors (ECOG performance status, TKI chosen, Race and EGFR status) used at randomization will also be summarized by treatment group and overall. In addition, summary statistics for tobacco history and number of pack years will also be presented. This will be done for SAF, FAS, PPS, PKAS and PDAS by treatment group and overall.

Frequency tabulations for NSCLC disease history (primary diagnosis, most recent NSCLC stage, primary tumor stage, regional lymph node stage and distant metastasis) as well as the descriptive statistics for duration of disease (duration since initial diagnosis and duration since date of locally advanced or metastatic disease diagnosis) will be provided by treatment group and overall for FAS and SAF.

EGFR and T790 status (exon 19 deletion, exon 21 L858R and T790M mutation status) from local and central assessments will be summarized separately by treatment group and overall for FAS and SAF.

Medical history other than NSCLC and conditions existing at Baseline will be coded in MedDRA and summarized by System Organ Class (SOC) and Preferred Term (PT) as well as by PT alone, by treatment group and overall for the FAS. Baseline conditions are defined as those ongoing at the time of informed consent or arise following the time of informed

consent and before the first dose of study drug. For ongoing medical conditions, Common Terminology Criteria for Adverse Events (CTCAE) grade will be provided in listing.

7.2.4 Previous and Concomitant Medications

Previous medications are coded with WHO-DD, and will be summarized by therapeutic subgroup (ATC 2nd level) ,chemical subgroup (ATC 4th level) , preferred WHO name and route for the SAF.

As with previous medication, concomitant medication will be summarized for each treatment group by therapeutic subgroup (ATC 2nd level) ,chemical subgroup (ATC 4th level) ,preferred WHO name and route for the SAF. Subjects taking the same medication multiple times will be counted once per medication and investigational period. All previous and concomitant medications will be coded by indication-specific ATC.

7.2.5 Prior Anti-Cancer Drug Therapy for NSCLC

Frequency tabulations of subjects with prior NSCLC drug therapy and best response to prior drug therapy will be presented by treatment group and overall for FAS.

7.2.6 Prior Radiation Therapy

Frequency tabulations of subjects with prior radiation therapy will be presented by treatment group and overall for FAS.

7.2.7 Prior Cancer Surgical Procedures

Frequency tabulations of subjects with prior cancer surgical procedures will be presented by treatment group and overall for FAS.

7.2.8 Previous and Concomitant Non-Medication Therapy

Frequency tabulations of subjects with previous and non-medication therapy and its reason for use will be presented by treatment group and overall for SAF.

7.3 Study Drugs

7.3.1 Exposure

The following information on drug exposure will be presented for each treatment regimen for the SAF:

- Descriptive statistics for cumulative amount of the drug subject was exposed to and actual dose intensity; and
- Number and percent of subject with dose adjustments or interruptions and reasons for dose adjustments and interruptions by treatment regimen.

7.3.2 Relative Dose Intensity

RDI will be examined for subjects in the SAF whose total study drug count and first and last days of treatment are known.

RDI will be summarized in two ways for the SAF:

- Descriptive statistics will be presented by treatment regimen.
- RDI will be categorized according to the following categories by treatment regimen:
    - less than 50%
    - at least 50%, less or equal to 80%
    - greater than 80%
    - Unknown.

7.4 Analysis of Efficacy

Efficacy analyses will be conducted on the FAS and PPS. The interpretation of results from statistical tests will be based on the FAS. The PPS will be used to assess the robustness of the results from the statistical tests based on the FAS. All randomized subjects will be analyzed according to the treatment to which they are randomized.

7.4.1 Analysis of Primary Endpoint

The primary analysis will be performed on progression-free survival as assed by the IRR (PFS#1) on the FAS. In order to compare the PFS#1 between ASP8273 and the comparator Erlotinib/Gefitinib, the null hypothesis will be constructed:

- H_{01}: PFS of ASP8273 is not better than the comparator erlotinib/gefitinib

The accompanying alternative hypothesis is:

- H_{11}: PFS of ASP8273 is better than the comparator erlotinib/gefitinib

The primary analysis will be performed when approximately 362 PFS#1 events have been observed. Comparison of ASP8273 and erlotinib/gefitinib will be tested at 1-sided significance level of 0.025 .

The distribution of PFS will be estimated for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using log-rank test stratified by ECOG PS (0 and 1 combined versus 2), EGFR mutation type (exon 19 vs exon 21), and TKI chosen (erlotinib versus gefitinib). In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

The sensitivity analyses for PFS include:

- The primary analysis repeated for PPS
- The primary analysis including clinical progression as an event may be performed on the FAS
- If there is a difference in the stratification values between IRT and the CRF, primary analysis may be repeated on the FAS using CRF stratification values

- PFS#1 as assessed by the IRR will be analyzed using restricted mean survival time (RMST) approach from time 0 to 36 months. The restricted mean is a measure of average survival from time 0 to a specified timepoint. This approach does not require any model assumptions such as proportional hazards.

Table 4 PFS#1 Definition

Situation	Date of Event or Censor	Outcome
No evaluable post-baseline imaging assessments, no death	Date of randomization	Censor
Subject did not receive new anti-cancer therapy:		
Radiographical progression documented per RECIST v1.1	Date of radiological PD	Event
No radiographical progression, but death recorded on eCRF	Date of death	Event
No radiographical progression nor death	Date of last radiological assessment or at the date of randomization if no post-baseline radiological assessment is available	Censor
Subject received new anti-cancer (ACT) therapy:		
Radiographical progression documented per RECIST v1.1 after new ACT	Date of last radiological assessment before start of new anti-cancer therapy or at the date of randomization if no post-baseline radiological assessment is available	Censor
Radiographical progression documented per RECIST v1.1 before new ACT	Date of radiological PD	Event
No radiographical progression before new ACT but death recorded	Date of last radiological assessment before start of new anti-cancer therapy or at the date of randomization if no post-baseline radiological assessment is available	Censor
No radiographical progression nor death	Date of last radiological assessment before start of new anti-cancer therapy or at the date of randomization if no post-baseline radiological assessment is available	Censor
Missed 2 scheduled radiological assessments:		
If radiographical progression or death occurs after missing 2 scheduled radiological assessments	Date of last radiological assessment or at the date of randomization if no post-baseline radiological assessment is available	Censor

PFS#1 = Date of Event or Censor – Date of Randomization +1

In addition, since it is likely that K-M curves of PFS#1 as assessed by IRR will separate late in the study, Harrington-Fleming test will be performed as exploratory analysis with pre-specified weight of p=0 and q=0.5. Different weights may be used as well.

To investigate non-proportionality between the treatment arms, the LOG(–LOG) survivor function versus time curves will be plotted. The scaled Schoenfeld residuals by time plot will be examined for evidence of a non-zero correlation, which indicates non-proportionality. Piecewise Cox models may be explored given evidence of non-proportional hazards.

Hochberg procedure will be used to control the overall error rate at 1-sided 0.025 level for the secondary efficacy endpoints (OS and ORR). Only when the primary endpoint significantly favors the ASP8273 arm will secondary endpoints be tested.

7.4.1.1 Overall Survival

The distribution of OS will be estimated for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using the log-rank test stratified by ECOG PS (0 and 1 combined versus 2), EGFR mutation type (exon 19 deletion or mutations in exon 21 [L858R]) and TKI chosen (erlotinib versus gefitinib). In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

Comparison of ASP8273 and erlotinib/gefitinib will be tested at 1-sided significance level of 0.025. The null hypothesis will be:

- H_{02}: OS of ASP8273 is not better than the comparator erlotinib/gefitinib

The accompanying alternative hypothesis is:

- H_{12}: OS of ASP8273 is better than the comparator erlotinib/gefitinib

OS will be evaluated on the FAS and PPS. Additional sensitivity analysis for OS will include:

- If there is a difference in the stratification values between IRT and the CRF, primary analysis may be repeated on the FAS using CRF stratification values

Table 5 OS Definition

Situation	Date of Event or Censor	Outcome
Death before analysis cutoff date	Date of death	Event
Death after analysis cutoff date	Analysis cutoff date	Censor
Last known alive is before cutoff date	Last known alive date	Censor
Last known alive is after cutoff date	Analysis cutoff date	Censor

OS = Date of Event or Censor – Date of Randomization +1

To investigate non-proportionality between the treatment arms, the analyses for PFS#1 by IRR will be carried out for OS.

Subsequent anti-cancer therapies will be collected and summarized by treatment arm within drug class. Imbalance in the use of subsequent anti-cancer therapy may potential bias the inference in OS. The following methods may be used to assess the impact of the use of subsequent anti-cancer therapy in OS.

- Censor OS at the last contact date prior to the initiation of subsequent anti-cancer therapy
- Rank-preserving structural failure time method (RPSFT) to correct for confounding introduced by the change of treatment. The RPSFT model is based on an accelerated failure time model and uses a structural assumption of time-proportionality instead of a proportional hazards assumption as used in the Cox model
- Inverse probability of censoring weights (IPCW).

7.4.1.2 Objective Response Rate

The comparison of ORR without confirmation as assessed by the blinded IRR between Arm A and Arm B will be performed using stratified CMH test. In addition, ORR for each arm will be estimated and corresponding 95% confidence interval will be constructed.

ORR will be assessed on the FAS.

Comparison of ASP8273 and erlotinib/gefitinib will be tested at 1-sided significance level of 0.025. The null hypothesis will be:

- H_{03}: ORR of ASP8273 is not better than the comparator erlotinib/gefitinib

The accompanying alternative hypothesis is:

- H_{13}: ORR of ASP8273 is better than the comparator erlotinib/gefitinib

Sensitivity analyses for ORR include:

- Repeat the analysis on the PPS.
- Repeat the analysis for ORR with confirmation as assessed by the blinded IRR
- Repeat the analysis for ORR without confirmation as assessed by the local investigator

7.4.1.3 Disease Control Rate

The comparison of DCR without confirmation as assessed by the blinded IRR between Arm A and Arm B will be performed using stratified CMH test. In addition, DCR for each arm will be estimated and corresponding 95% confidence interval will be constructed.

DCR will be assessed on the FAS.

Sensitivity analyses for DCR include:

- Repeat the analysis on the PPS.
- Repeat the analysis for DCR with confirmation as assessed by the blinded IRR
- Repeat the analysis for DCR without confirmation as assessed by the local investigator

7.4.1.4 PFS#1 by local investigator

The distribution of PFS#1 by local investigator will be estimated on the FAS for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using stratified log-rank test as specified in [Section 7.4.1]. In addition, stratified Cox

proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

7.4.1.5 Duration of Response

The distribution of DOR will be estimated on the FAS for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using stratified log-rank test as specified in [Section 7.4.1]. In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

Table 6 DOR Definition

Situation	Date of Event or Censor	Outcome
Subject with CR or PR (no confirmation required):		
Radiographical progression documented per RECIST v1.1	Date of radiological PD	Event
No radiographical progression	Date of last radiological assessment or at the date of first CR or PR if no post-baseline radiological assessment is available	Censor

DOR = Date of Event or Censor – Date of first CR or PR +1

7.4.2 Analysis of Exploratory Endpoint

7.4.2.1 PFS#2

The distribution of PFS#2 will be estimated on the FAS for each treatment arm using Kaplan-Meier methodology and compared between Arm A and Arm B using stratified log-rank test as specified in [Section 7.4.1]. In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval.

Table 7 PFS#2 Definition

Situation	Date of Event or Censor	Outcome
Subject did not receive new anti-cancer therapy	Day 1	Censor
Subject received new anti-cancer therapy:		
Subject has investigator-determined progression on that treatment or ended that treatment or started another therapy or death	Minimum of [Date of investigator determined PD, date of death, new ACT end date, or start date of another therapy]	Event
Subject did not die nor progressed nor stopped the new ACT nor started another therapy at the time of analysis cutoff date	Last known alive date	Censor

PFS#2 = Date of Event or Censor – Date of initiation of new systemic anti-cancer treatment +1

7.4.3 QoL/PRO

Descriptive QoL and PRO analyses will be performed on the FAS. Completion rate for each questionnaire will be summarized. Additional analyses will be discussed in details in a separate PRO SAP.

7.4.3.1 FACT-EGFRI-18

Means and standard deviations and change from baseline at each scheduled assessment will be reported for each of the FACT-EGFRI-18 domains (physical, social/emotional and functional well-being). The analyses will include data from the baseline assessment through the last available data.

Detailed scoring instructions for these domains are provided in Appendix 10.3. All time point data will be included and summarized.

7.4.3.2 EORTC-QLQ-C30

Means and standard deviations and change from baseline at each scheduled assessment will be reported for each of the QLQ-C30 subscales. The analyses will include data from the baseline assessment through the last available data.

Detailed scoring instructions for these scales are provided in Appendix 10.5. All time point data will be included and summarized.

7.4.3.3 EORTC-QLQ-LC13

Means and standard deviations and change from baseline at each scheduled assessment will be reported for each of the QLQ-LC13 subscales. The analyses will include data from the baseline assessment through the last available data.

Detailed scoring instructions for these scales are provided in Appendix 10.4. All time point data will be included and summarized.

7.4.3.4 EQ-5D-5L

Descriptive characteristics of EQ-5D-5L will be summarized from baseline assessment through the last available data. Frequency and the percentage of reported problems for each level for each dimension will be provided. Shift table showing shift in each dimension score from baseline to each post-baseline visit will be provided for the 5 dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). 2D bar plots will be generated to summarize the reported problems for each level for each dimension.

EQ-5D-5L VAS will be summarized by treatment group at each visit using descriptive statistics (mean, standard deviation, minimum, maximum and median). Additionally, a within-subject change will be calculated as the post-baseline score minus the baseline score and summarized in the same way. ANOVA model will be used to evaluate change from baseline to post-baseline visits for the EQ-5D-5L VAS including treatment as a fixed factor, and baseline score as a covariate.

Detailed scoring instructions for these dimensions are provided in Appendix 10.6. All time point data will be included and summarized.

7.5 Analysis of Safety

All analysis of safety will be presented by treatment group and overall for SAF, unless specified otherwise.

7.5.1 Adverse Events

For the purpose of safety assessments in this study, events recorded during the pre-investigational period will be classified as Baseline Signs and Symptoms. All adverse event (AE) recorded on treatment including within 30 days from the last study treatment will be summarized.

The coding dictionary for this study will be MedDRA v18.0. It will be used to summarize AEs by SOC and PT. AEs will be graded using National Cancer Institute's Common Terminology Criteria for AEs (NCI-CTCAE).

An overview table will include the following details by treatment group and overall:

- Number and percentage of subjects with TEAEs,
- Number and percentage of subjects with causally drug-related TEAEs,
- Number and percentage of subjects with serious TEAEs and Astellas upgraded serious TEAE,
- Number and percentage of subjects with serious drug-related TEAEs and Astellas upgraded serious drug-related TEAE,
- Number and percentage of subjects with TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with drug-related TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with grade 3 or higher TEAE
- Number and percentage of subjects with TEAEs leading to dose reduction,
- Number and percentage of subjects with drug-related TEAEs leading to dose reduction,
- Number and percentage of subjects with TEAEs leading to drug interruption,
- Number and percentage of subjects with drug-related TEAEs leading to drug interruption, and
- Number of deaths.

The number and percentage of subjects with TEAEs, as classified by SOC and PT will be summarized by treatment group and overall. Summaries will be provided for:

- TEAEs
- Drug-related TEAEs,
- serious TEAEs and Astellas upgraded serious TEAE,
- drug-related serious TEAEs and drug-related Astellas upgraded serious TEAE,
- TEAEs leading to permanent discontinuation of study drug,
- Drug-related TEAEs leading to permanent discontinuation of study drug,
- TEAEs leading to dose reduction,
- Drug-related TEAEs leading to dose reduction,
- TEAEs leading to drug interruption,
- Drug-related TEAEs leading to drug interruption,
- TEAEs excluding serious adverse events that equal to or exceed a threshold of 10% in any dose level, and
- Common TEAEs that equal to or exceed a threshold of 10% in any dose level.
- TEAE with CTCAE Grade 3 or higher

The number and percentage of subjects with TEAEs and TEAEs leading to death, as classified by PT only, will be summarized by treatment group and overall.

AE summary tables will include subject counts as opposed to AE counts. If a subject experiences more than one episode of a particular AE, that subject will be counted only once for that event. If a subject has more than one AE that code to the same preferred term, the subject will be counted only once for that preferred term. Similarly, if a subject has more than one AE within a body system, the subject will be counted only once in that body system.

TEAEs and the number and percentage of subjects with TEAEs, as classified by SOC and PT will also be summarized by NCI-CTCAE severity grade and by relationship to study drug. If an adverse event changes in severity grade or relationship, then the subject will be counted only once with the worst severity grade and highest degree of relationship. The adverse event however will be presented in each category they were classified to. If a subject has an event more than once with missing severity grade/relationship and with non-missing severity grade, then the subject will be counted as the highest non-missing grade. If a subject has an event more than once with missing relationship and with non-missing relationship, then the subject will be counted as missing. Drug-related TEAEs will be presented in a similar way by severity only.

The exposure adjusted subject incidence rates of adverse events may be compared between treatment groups for selected categories of adverse events. The exposure adjusted subject incidence (subjects/subject-year) is defined for any particular adverse event as the number of subjects in a treatment group that experienced the specific event at least once divided by the total amount of time (100 subject-year) the subjects in the treatment group have been exposed

to the study drug. For subjects that experience the event, their exposure time will be calculated as the time between the first dose date up to the initial occurrence of the event. For subjects who do not experience the event, their exposure time will be calculated as the time between the first dose date through 30 days after the last dose date, the date of death, or the end of study date, whichever is earlier. For subjects that experience multiple events, their exposure time will be calculated up to the onset of the first event.

The number and percentage of subjects with treatment-emergent adverse events of interest, as classified by SOC and PT will also be summarized by treatment group and overall. The list of adverse events of interest to be summarized may change during the course of the study due to ongoing pharmacovigilance. It will be finalized before the database hard lock.

All AEs, deaths, SAEs and withdrawals due to adverse events will be displayed in listings.

All AEs occurring during or after the subject has discontinued the study are to be followed up until resolved or judged to be no longer clinically significant, or until they become chronic to the extent that they can be fully characterized.

7.5.2 Clinical Laboratory Evaluation

The baseline visit is the last measurement taken prior to initial study drug administration.

Selected quantitative clinical laboratory variables, i.e. hematology, biochemistry, coagulation and urinalysis will be summarized using mean, standard deviation, minimum, maximum and median by treatment group at each visit. Additionally, a within-subject change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Changes in laboratory values will be summarized for baseline values versus minimum, maximum and last post-baseline values. Plots of median lab values at each scheduled assessment time by treatment group may be provided for each laboratory parameter.

Frequency tabulations of selected qualitative clinical laboratory variables (i.e. urinalysis) will be presented by treatment group at each visit.

Laboratory results will also be graded using NCI-CTCAE, where possible. Parameters that have criteria available for both low and high values, i.e., hypo- and hyper-, will be summarized for both criteria. The same subject can be counted for both values if the subject has different laboratory values meeting each criterion. NCI-CTCAE grade of laboratory evaluations will be summarized by number and percentage of subjects for each visit. Shift tables of NCI-CTCAE grade change from baseline to worst post-baseline grade for selected laboratory parameters will also be presented.

The list of laboratory parameters to be summarized may change during the course of the study due to ongoing pharmacovigilance. It will be finalized before the database hard lock.

Laboratory data will be displayed in listings.

7.5.2.1 Liver Enzymes and Total Bilirubin

The following potentially clinically significant criteria in liver function tests for Alkaline Phosphatase (ALP), Alanine Transaminase (ALT), total bilirubin, Aspartate Transaminase (AST) and their combination are defined. The subject's highest value during the investigational period will be used.

Parameter	Criteria
ALT	> 3xULN
	> 5xULN
	> 8xULN
AST	> 3xULN
	> 5xULN
	> 8xULN
ALT or AST	> 3xULN
	> 5xULN
	> 8xULN
Total Bilirubin	> 2xULN
ALP	> 1.5xULN
ALT and/or AST AND Total Bilirubin(*)	(ALT and/or AST > 3xULN) and total bilirubin >= 2xULN

(*) Combination of values measured within same sample

The number and percentage of subjects with potentially clinically significant values in liver enzymes and total bilirubin during the investigational period will be presented by treatment group.

7.5.3 Vital Signs

The baseline visit is the last measurement taken prior to initial study drug administration.

Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate and body temperature) will be summarized using mean, standard deviation, minimum, maximum and median by treatment group and visit. Additionally, a within-subject change will be calculated per visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group and visit.

7.5.4 Electrocardiograms (ECGs)

Number and percentage of subjects with normal and abnormal results as assessed locally for the overall interpretation will be tabulated by treatment group at each treatment visit and time point.

7.5.5 Pregnancies

A detailed listing of all pregnancies will be provided.

7.5.6 Physical Examinations

A listing of physical examination results will be produced.

7.5.7 ECOG Performance Status

Summary statistics (number and percent of subjects) for each category of the ECOG performance status at each assessment will be provided. The change from baseline to final visit or early termination will also be summarized. Negative change scores indicate an improvement. Positive scores indicate a decline in performance.

Time to deterioration from baseline in ECOG performance status defined as the change in score from 0 to 2 (or worse) or 1 to 2 (or worse) or 2 to 3 (or worse) will also be summarized. The time to deterioration will be calculated in reference to the randomization date.

7.6 Analysis of PK

PK analysis will be conducted on the PKAS.

Plasma concentrations of ASP8273 and its metabolite(s) (if applicable) will be summarized for Arm A (ASP8273 arm) and where appropriate by nominal time points using descriptive statistics, including number of subjects, mean, standard deviation, minimum, median, maximum, geometric mean, and coefficient of variation (CV).

7.7 Analysis of Biomarker

Selected efficacy endpoints will be summarized for the following groups:

- EGFR mutation detected by local and central assessments
- EGFR mutation detected by central assessment but not local

Additional potential biomarkers that may affect treatment outcome will be explored and analyses details will be defined in a separate biomarker analysis plan.

7.8 Subgroups of Interest

PFS#1 by IRR, OS, DOR and ORR by IRR will be summarized for the subgroups defined on the basis of the categorized variables listed below:

Grouping variable	Subgroups
Age group	< 65 years
	>=65 years - <75years
	>=75 years
Sex	Female
	Male
Race	Asian
	Non-Asian
ECOG status at baseline	0 and 1
	2
EGFR status	Exon 19 Deletion
	Exon 21 L858R
TKI chosen	Erlotinib
	Gefitinib

Additional analyses may be performed to explore treatment outcome between ECOG status at baseline scores 0 and 1. Forest plots for PFS#1 by IRR and OS will be produced to summarized the treatment effect across subgroups. PFS#1 by IRR, OS, DOR and ORR by IRR may also be summarized using three treatment arms (i.e. ASP8273, Erlotinib and Gefitinib).

7.9 Interim Analysis (and Early Discontinuation of the Clinical Study)

An interim futility analysis is planned to occur after 210 PFS events are observed, which is expected to occur approximately 22 months after the first subject is randomized. If the observed HR is less than 1.1, recruitment into the study will continue as planned until 600 subjects are recruited. If the observed HR at the interim is larger than 1.1, recruitment into the study may be stopped at the interim analysis. Further review of the efficacy data, including secondary endpoints, will be performed by the sponsor to determine whether other evidence of clinical benefit justifies reopening enrollment. If the data do not justify reopening enrollment, enrollment may remain closed, and a decision will be made by the sponsor regarding termination of the study. If the futility analysis takes place after enrollment is completed and the HR is greater than 1.1, subjects who in the opinion of the investigator and the sponsor's Medical Monitor are continuing to derive benefit may continue to receive study drug as randomized. As the interim analysis is for futility only, type I error at final analysis will not be affected.

The interim analysis will be conducted by the IDMC. In addition safety data reviews during the trial will be conducted by the IDMC on a periodic basis. For example, the IDMC will review safety data after the first 50 patients have been randomized plus 3 months. The full procedures for IDMC safety review will be described in a separate IDMC charter.

7.10 Handling of Missing Data, Outliers, Visit Windows, and Other Information

7.10.1 Missing Data

Every effort will be made to resolve missing or incomplete dates for adverse events and concomitant medications. If a partial date cannot be resolved, the most conservative imputation methods will be used to complete the missing information.

Missing or partial start and stop dates of adverse events, non-medication therapy and concomitant medication will be imputed using the following algorithm:

- Imputation rules for partial or missing stop dates:
  - If the month and year are present, then impute as the last day of that month.
  - If only the year is present, impute as December 31 of that year.
  - If the stop date is entirely missing, assume the event or medication is ongoing.
- Imputation rules for partial or missing start dates:

Start Date		Stop Date						
		Complete: *yyyymmdd*		Partial: *yyyymm*		Partial: *yyyy*		Missing
		< 1st dose	≥ 1st dose	< 1st dose *yyyymm*	≥ 1st dose *yyyymm*	< 1st dose *yyyy*	≥ 1st dose *yyyy*	
Partial: *yyyymm*	= 1st dose *yyyymm*	2	1	2	1	n/a	1	1
	≠ 1st dose *yyyymm*		2		2	2	2	2
Partial: *yyyy*	= 1st dose *yyyy*	3	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute as the date of first dose; 2 = Impute as the first of the month; 3 = Impute as January 1 of the year; 4 = Impute as January 1 of the stop year

The imputed dates will be used to determine whether an AE is/is not treatment emergent. Listings of AEs, non-medication therapy and concomitant medications will present the actual partial dates; imputed dates will not be shown.

For secondary endpoint OS, missing or incomplete death date will be imputed as the earliest feasible date on or after the date last known to be alive as the examples shown in the table below.

Incomplete Date of Death (YYYY MMM DD)	Date Last Know to be Alive (YYYY MMM DD)	Imputed Date of Death (YYYY MMM DD)
2005 APR ??	2005 MAR 31	2005 APR 01
2005 ??? 13	2005 MAR 31	2005 APR 13
2005 ??? ??	2005 MAR 31	2005 MAR 31
???? APR ??	2005 MAR 31	2005 APR 01
???? APR 13	2005 MAR 31	2005 APR 13
???? ??? ??	2005 MAR 31	2005 MAR 31

7.10.2 Outliers

All values will be included in the analyses.

7.10.3 Visit Windows

Not applicable. Nominal visits will be used in the by visit summary. Values from unscheduled visits will be included in the summary of extreme cases (e.g. summary of worst post-baseline, summary of minimum post-baseline, summary of maximum post-baseline). For efficacy endpoints, all values (scheduled and unscheduled) will be included in the analysis.

7.10.4 Blinding

Although the study is an open label study, to increase the credibility of study results, the sponsor statistician's access to the randomized treatment assignment information will be limited. This will help reduce potential bias due to the sponsor knowing the treatment effect due to unintentional efficacy and safety summary by treatment. The following procedures will be implemented in this study:

- The randomized treatment code will not be transferred to statistical programmers, study statistician and supporting statisticians before the database lock. Datasets with scrambled treatment code will be transferred to prepare analysis programs.
- The study statistician, supporting statisticians and statistical programmers will have no direct access to the randomized treatment information before the database lock.
- Study managers and other study team members may have the access to the treatment information at the individual subject level.
- No treatment difference will be summarized during the study, except planned interim analysis or by IDMC request.

8 DOCUMENT REVISION HISTORY

Version	Date	Changes	Comment/rationale for change
1.0	02-Oct-2015	NA	Document finalized
2.0	24-Jun-2016	Section 2: Flow chart Table 1 and Table 2: Schedule of assessment Sections 3.1.2, 3.1.3 and 3.2.1 - Quality of Life (QoL) was moved from an exploratory to a secondary endpoint. Section 4 - Sample size and number of events required was changed to 600 and 420, respectively. Section 7.10 - Updated the sample size and the number of events required for the interim analysis	Changes were made to align the SAP with the protocol version 2.0
		Section 6.6 Provided additional text on how to handle duration variables when the date is incomplete or missing Added definition of previous and concomitant non-medication therapy	To clarify the rule in duration derivation when dates are incomplete or missing and also clarify the definition of previous and concomitant non-medication therapy.
		Section 7.1 Provided additional text in derivation of study day and baseline	To provide guidance on how to derive study day and baseline since different reference date will be used for safety and efficacy assessments
		Section 7.4.1 Added sensitivity analysis and exploratory analysis to the primary endpoint (PFS#1 as assessed by IRR)	Late separation of survival curves is expected to occur
		Section 7.4.1.1 Added sensitivity analysis and exploratory analysis to overall survival	To investigate the potential effect of subsequent therapies on overall survival
		Section 7.5.1 Added additional summaries for adverse events	Adverse events leading to dose reduction and drug interruptions are important safety information
		Section 7.5.7 Added text in time to deterioration derivation	Provided clarification regarding the reference date to be used for time to deterioration

Version	**Date**	**Changes**	**Comment/rationale for change**
		Section 7.2.4 Added a text to clarify the coding process for concomitant medication	Concomitant medications will be coded by indication-specific ATC
		Section 7.2.8 New section for previous and concomitant non-medication therapy	Added section on how to summarize previous and concomitant non-medication therapy

9 REFERENCES

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez N, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. Mar 3 1993;85(5):365-76.

Alan Agresti (1990). Categorical data analysis. New York: Wiley.

Bezjak A, Lee CW, Ding K, Brundage M, Winton T, Graham B, et al. Quality-of-life outcomes for adjuvant chemotherapy in early-stage non-small-cell lung cancer: results from a randomized trial, JBR.10. J Clin Oncol. Nov 1 2008;26(31):5052-9.

Clopper CJ, Pearson, ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, Biometrika, 1934;26(4):404-13.

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz D, Ford R, Dancey J, et al. New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228-47.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC, 2001.

Grunfeld E, Watters JM, Urquhart R, O'Rourke K, Jaffey J, Maziak DE, et al. A prospective study of peri-diagnostic and surgical wait times for patients with presumptive colorectal, lung, or prostate cancer. Br J Cancer. Jan 13 2009;100(1):56-62.

Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 1988;75(4):800-2.

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

Lis CG, Rodeghier M, Grutsch JF, Gupta D. Distribution and determinants of patient satisfaction in oncology with a focus on health related quality of life. BMC Health Serv Res. 2009;9:190.

Maringwa JT, Quinten C, King M, Ringash J, Osoba D, Coens C, et al. Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. Support Care Cancer. Nov 2011;19(11):1753-60.

Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Seto T, et al. Updated overall survival results of WJTOG 3405, a randomized phase III trial comparing gefitinib (G) with cisplatin plus docetaxel (CD) as the first-line treatment for patients with non-small cell lung cancer harboring mutations of the epidermal growth factor receptor (EGFR). J Clin Oncol. 2012;30(suppl;abstr 7521).

Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol. 2010;11:121-8.

O'Brien ME, Anderson H, Kaukel E, O'Byrne K, Pawlicki M, Von Pawel J, et al. SRL172 (killed Mycobacterium vaccae) in addition to standard chemotherapy improves quality of life without affecting survival, in patients with advanced non-small-cell lung cancer: phase III results. Ann Oncol. Jun 2004;15(6):906-14.

Pickard AS, Wilke CT, Lin HW, Lloyd A. Health Utilities Using the EQ-5D in Studies of Cancer. PharmacoEconomics 2007;25:365–84

Wagner LI, Berg SR, Gandhi M, Hlubocky FJ, Webster K, Aneja M, et al. ME. The development of a Functional Assessment of Cancer Therapy (FACT) questionnaire to assess dermatologic symptoms associated with epidermal growth factor receptor inhibitors (FACT-EGFRI-18). Support Care Cancer. Apr 2013;21(4):1033-41

Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol. 2011;12:735-42.

10 APPENDICES

10.1 Appendix 1: Key Contributors and Approvers

List of Key Contributors and Approvers

Key Contributors

The following contributed to or reviewed this Statistical Analysis Plan as relevant to their indicated discipline or role.

Primary author

[redacted], Ph.D. [redacted]

Contributors and Reviewers

[redacted] Ph.D. [redacted]

[redacted], PharmD [redacted]

[redacted], MD [redacted]

[redacted], M.S. [redacted]

[redacted], Ph.D. [redacted]

[redacted] [redacted]

[redacted], Ph.D.

[redacted]

[redacted], MS

[redacted]

[redacted]

[redacted]

[redacted], Pharm.D.

[redacted]

[redacted]

[redacted]

[redacted]

Author and Approver Signatories

(E-signatures are attached at end of document)

[redacted]/Ph.D., [redacted] was the study statistician for this study and the primary author of this Statistical Analysis Plan

[redacted]/Ph.D., [redacted] was the biostatistics peer reviewer of this Statistical Analysis Plan

This Statistical Analysis Plan was approved by:

[redacted], MD

[redacted], Ph.D.

[redacted], MS

10.2 Appendix 2: Blinding Process

As specified in SAP Section 7.10.4, although the study is an open label study, to increase the credibility of study results, the sponsor statistician's access to the randomized treatment assignment information in study database will be limited. This will reduce potential bias due to the sponsor knowing the treatment effect due to unintentional efficacy and safety summary by treatment.

For the purpose of controlling unintentional assessment of efficacy and safety summary by treatment, the study statisticians and programmers will have no access to treatment codes until the time of final database lock (DBL). For this study, treatment codes can be obtained from two sources: IRT vendor and EDC system. The following procedures should be applied when handling data extraction, data transferring and TLFs preparation to maintain the blinding during the entire study conducting period before the final DBL.

- The treatment codes collected in EDC system will not be extracted until the final DBL. The treatment variable will be blocked during any data extraction before the final DBL. The option for blocking treatment variable will be set up before the first data extraction.
- Dummy treatment codes will be created by study programmers for preparing the SDTM, ADaM and TLFs for interim analysis validation and final data analysis.
- Raw data (SDTM/ADAM if needed) without treatment codes (or with dummy codes) will be provided by the study team to IDAC to conduct interim analysis or other DMC requested analysis.
- Separate IRT transfer requests will be sent directly by independent statistician to IRT vendor for transferring treatment codes to IDAC during the formal interim analysis or other DMC requested analysis.
- PK data will be transferred upon request during interim analysis and will be transferred directly from PK vendor to IDAC.

Study manager and other study team members may have the access to the treatment assignment information at the individual subject level. Dosing or other information which may relate to treatment assignment will not be blinded at the individual subject level due to operational considerations. Efficacy and safety information will not be summarized by treatment during the study, except at the planned interim analysis or other DMC requested analysis conducted by IDAC.

10.3 Appendix 3: FACT-EGFR-18 Scoring Algorithm

Functional Assessment of Cancer Therapy – EGFR Inhibitors (FACT-EGFRI 18)

Scoring Guidelines

Instructions:*
1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the symptom index score.
4. As with all FACIT questionnaires, a high score is good. Therefore, a score of "0" is a severely symptomatic patient and the highest possible score is an asymptomatic patient.

Scale	Item Code	Reverse item?		Item response	Item Score
FACT-EGFRI	ST4	4	-	_____	=_____
	ST5	4	-	_____	=_____
	ST6	4	-	_____	=_____
Score range: 0-72	ST7	4	-	_____	=_____
	ST9	4	-	_____	=_____
	ST32	4	-	_____	=_____
	ST22	4	-	_____	=_____
	ST17	4	-	_____	=_____
	ST24	4	-	_____	=_____
	ST37	4	-	_____	=_____
	ST26	4	-	_____	=_____
	ST34	4	-	_____	=_____
	ST38	4	-	_____	=_____
	ST16	4	-	_____	=_____
	ST15	4	-	_____	=_____
	ST29	4	-	_____	=_____
	B5	4	-	_____	=_____
	ST11	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 18: _____
Divide by number of items answered: _____ **= FACT-EGFRI score**

10.4 Appendix 4: EORTC QLQ-LC13 Scoring Algorithm

Lung cancer module: QLQ-LC13

The lung cancer module is meant for use among a wide range of lung cancer patients varying in disease stage and treatment modality (Bergman *et al.*, 1994). The module comprises 13 questions (Appendix 2c). This module was constructed in parallel with the core QLQ-C30, before the guidelines on module development had been established. It was field tested together with the previous versions of the core questionnaire (QLQ-C36, QLQ-C30(V1)). The module is designed for use among patients receiving treatment with chemotherapy and / or radiotherapy. The QLQ-LC13 includes questions assessing lung cancer-associated symptoms (cough, haemoptysis, dyspnoea and site specific pain), treatment-related side effects (sore mouth, dysphagia, peripheral neuropathy and alopecia) and pain medication. The module was field tested together with the previous versions of the core questionnaire.

Scoring of the lung cancer module

The lung cancer module incorporates one multi-item scale to assess dyspnoea, and a series of single items assessing pain, coughing, sore mouth, dysphagia, peripheral neuropathy, alopecia, and haemoptysis.

The scoring approach for the QLQ-LC13 is identical in principle to that for the symptom scales / single items of the QLQ-C30.

Scale name	Scale	Number of items	Item range*	QLQ-LC13 Item numbers	†
Symptom scales / items					
Dyspnoea†	LCDY	3†	3	3,4,5	X
Coughing	LCCO	1	3	1	
Haemoptysis	LCHA	1	3	2	
Sore mouth	LCSM	1	3	6	
Dysphagia	LCDS	1	3	7	
Peripheral neuropathy	LCPN	1	3	8	
Alopecia	LCHR	1	3	9	
Pain in chest	LCPC	1	3	10	
Pain in arm or shoulder	LCPA	1	3	11	
Pain in other parts	LCPO	1	3	12	

* "Item range" is the difference between the possible maximum and the minimum response to individual items.

† The dyspnoea scale should only be used if all three items have been answered. Some respondents ignore question 5 because they never climb stairs; in this case, the score for the dyspnoea scale would be biased if it were based upon the other two items. Hence if item 5 is missing then items 3 and 4 should be used as single-item measures.

10.5 Appendix 5: EORTC QLQ-30 Scoring Algorithm

Scoring the EORTC QLQ-C30 version 3.0

Table 1: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range*	**Version 3.0** Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS - 1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL:**

$$Score = \{(RS - 1)/range\} \times 100$$

Examples:

Emotional functioning

$$RawScore = (Q_{21} + Q_{22} + Q_{23} + Q_{24})/4$$
$$\text{EF } Score = \{1 - (RawScore - 1)/3\} \times 100$$

Fatigue

$$RawScore = (Q_{10} + Q_{12} + Q_{18})/3$$
$$\text{FA } Score = \{(RawScore - 1)/3\} \times 100$$

10.6 Appendix 6: EQ-5D-5L

2. Scoring the EQ-5D-5L descriptive system

The EQ-5D-5L descriptive system should be scored, for example, as follows:

Under each heading, please tick the ONE box that best describes your health TODAY

MOBILITY		
I have no problems in walking about	✓	✓
I have slight problems in walking about	☐	☐
I have moderate problems in walking about	☐	☐
I have severe problems in walking about	☐	☐ Level 1 is coded as a '1'
I am unable to walk about	☐	☐
SELF-CARE		
I have no problems washing or dressing myself	☐	☐
I have slight problems washing or dressing myself	✓	✓ Level 2 is coded as a '2'
I have moderate problems washing or dressing myself	☐	☐
I have severe problems washing or dressing myself	☐	☐
I am unable to wash or dress myself	☐	☐
USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*		
I have no problems doing my usual activities	☐	☐
I have slight problems doing my usual activities	☐	☐
I have moderate problems doing my usual activities	✓	✓ Level 3 is coded as a '3'
I have severe problems doing my usual activities	☐	☐
I am unable to do my usual activities	☐	☐
PAIN / DISCOMFORT		
I have no pain or discomfort	☐	☐
I have slight pain or discomfort	☐	☐
I have moderate pain or discomfort	☐	☐ Level 4 is coded as a '4'
I have severe pain or discomfort	✓	✓
I have extreme pain or discomfort	☐	☐
ANXIETY / DEPRESSION		
I am not anxious or depressed	☐	☐
I am slightly anxious or depressed	☐	☐
I am moderately anxious or depressed	☐	☐
I am severely anxious or depressed	☐	☐ Level 5 is coded as a '5'
I am extremely anxious or depressed	✓	✓

Levels of perceived problems are coded as follows:

This example identifies the health state '12345'.

***NB:** There should be only <u>ONE</u> response for each dimension*

***NB: Missing values** can be coded as '9'.*

***NB: Ambiguous values** (e.g. 2 boxes are ticked for a single dimension) should be treated as missing values.*