The following protocol information is provided solely to describe how the authors conducted the research underlying the published report associated with the following article:

PointBreak: A Randomized Phase III Study of Pemetrexed plus Carboplatin and Bevacizumab Followed by Maintenance Pemetrexed and Bevacizumab versus Paclitaxel plus Carboplatin and Bevacizumab Followed by Maintenance Bevacizumab in Patients with Stage IIIB or IV Nonsquamous Non-Small Cell Lung Cancer

Patel, et al

DOI: 10.1200/JCO.2012.47.9626

PAREXEL

Eli Lilly and Company

H3E-MC-JMHD

Randomized, Open-Label, Phase 3 Study of Pemetrexed plus Carboplatin and Bevacizumab Followed by Maintenance Pemetrexed and Bevacizumab versus Paclitaxel plus Carboplatin and Bevacizumab Followed by Maintenance Bevacizumab in Patients with Stage IIIB or IV Nonsquamous Non-Small Cell Lung Cancer

Statistical Analysis Plan

Version 2.2 [Final]
Date: 09Apr, 2012

TABLE OF CONTENTS

1 REVISION HISTORY 7
2 INTRODUCTION 9
3 STUDY OBJECTIVES 9
3.1 Primary Objective 9
3.2 Secondary Objectives 9
4 INVESTIGATIONAL PLAN 10
4.1 Overall Study Design and Plan 10
4.2 Efficacy and Safety Variables 11
4.2.1 Primary Efficacy Measure 11
4.2.2 Secondary Efficacy Measures 11
4.2.3 Health Outcome Measures 12
4.2.4 Safety Evaluations 13
4.2.5 Pharmacokinetics/Pharmacodynamics 13
5 STATISTICAL METHODS 13
5.1 General Considerations 13
5.2 Study Patients 14
5.2.1 Disposition of Patients 14
5.2.2 Important Protocol Violations 14
5.2.3 Analysis Populations 15
5.3 Demographic and Other Baseline Characteristics 15
5.3.1 Flagging of Concomitant Medications as Prior, Prior and Concomitant or Concomitant 16
5.4 Treatment Compliance 17
5.5 Efficacy Evaluation 17
5.5.1 Primary Efficacy Evaluation 17
5.5.2 Secondary Efficacy Evaluation 18
5.5.2.1 Progression Free Survival 18
5.5.2.2 Response Rate 18
5.5.2.3 Time to Progressive Disease 18
5.5.2.4 Disease Control Rate 18
5.5.3 Subgroup Analyses 18
5.5.3.1 Time-to-Event Analyses for Patients Who Receive Maintenance Therapy 18
5.5.3.2 Other Efficacy Subgroup Analyses 18
5.5.3.3 Other Safety Subgroup Analyses 19
5.5.4 Sensitivity Analyses for PFS, TTPD, and PFS without Grade 4 Toxicity (G4PFS) 19
5.6 Health Outcome Analyses 21
5.7 Safety Evaluation 21
5.7.1 Extent of Exposure 22
5.7.2 Adverse Events 22
5.7.3 Deaths, Serious Adverse Events, and Other Significant Adverse Events. 23
5.7.4 Clinical Laboratory Evaluation 23

5.7.5 Other Observations Related to Safety 24
5.7.6 Sensitivity Analyses for Adverse Events 24
5.8 Pharmacokinetic/Pharmacodynamic Analyses 25
5.8.1 Pharmacokinetic Parameter Estimation 25
5.8.2 Pharmacokinetic Statistical Inference 25
5.9 Translational Research 26
5.10 Patient Profiles 26
5.11 Determination of Sample Size 26
6 INTERIM ANALYSIS 27
7 REFERENCES 29
8 REPORTING OUTPUT 30

LIST OF ABBREVIATIONS

ADS	Analysis Data Set
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomic Therapeutic Class
AUC	Area Under The Curve
BOR	Best Overall Response
BSA	Body Surface Area
CI	Confidence Interval
CL	Systemic Clearance
Cmax	Maximum Concentration
CR	Complete Response
CRF	Case Report Form
CT	Computed Tomography
CTCAE	Common Terminology Criteria For Adverse Events
DCR	Disease Control Rate (CR + PR + SD)
DMC	Data Monitoring Committee
ECOG	Eastern Cooperative Oncology Group
EDS	Extract Data Set
FACT	Functional Assessment Of Cancer Therapy
G4PFS	PFS without grade 4 toxicity
GOG	Gynecologic Oncology Group
HR	Hazard Ratio
IVRS	Interactive Voice Response System
LCS	Lung Cancer Subscale
LS	Least Squares
MedDRA	Medical Dictionary of Regulatory Activities
MRI	Magnetic Resonance Imaging
NSCLC	Non-Small Cell Lung Cancer
NTx	Neurotoxicity
OS	Overall Survival
PFS	Progression-Free Survival
PK	Pharmacokinetic/Pharmacokinetics
PR	Partial Response
PT	Preferred Term
QOL	Quality of Life
RECIST	Response Criteria In Solid Tumors
RR	Response Rate
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
SOC	System Organ Class
T1/2	Half Life
TEAE	Treatment Emergent Adverse Event

TLF	Table, Listing, Figure
TOI	Trial Outcome Index
TTPD	Time To Progressive Disease
V	Apparent Volume of Distribution
Vss	Apparent Volume of Distribution at Steady State
WHO	World Health Organization

2 INTRODUCTION

This document presents the statistical analysis plan (SAP) for Eli Lilly and Company (Lilly) Protocol H3E-MC-JMHD: Randomized, Open-Label, Phase 3 Study of Pemetrexed plus Carboplatin and Bevacizumab Followed by Maintenance Pemetrexed and Bevacizumab versus Paclitaxel plus Carboplatin and Bevacizumab Followed by Maintenance Bevacizumab in Patients with Stage IIIB or IV Nonsquamous Non-Small Cell Lung Cancer (NSCLC). This analysis plan is based on the protocol amendment (d) approved on 02 Feb 2011 and contains definitions of analysis populations, derived variables, imputation rules for missing values, and statistical methods for the analysis of efficacy and safety parameters.

3 STUDY OBJECTIVES

3.1 Primary Objective

The primary objective is to compare overall survival (OS) for:

- Arm A: pemetrexed plus carboplatin plus bevacizumab, followed by maintenance bevacizumab plus pemetrexed
- Arm B: paclitaxel plus carboplatin plus bevacizumab, followed by maintenance bevacizumab

in the first-line induction and maintenance therapy for the treatment of patients with Stage III B (with pleural effusions) or IV nonsquamous NSCLC.

3.2 Secondary Objectives

The secondary objectives of the study are as follows:

Efficacy:

1. To compare the overall response rates (RRs) and the disease control rates (DCRs), assessed according to Response Evaluation Criteria in Solid Tumors (RECIST 1.0; Therasse et al. 2000), between the 2 treatment arms.

2. To compare the following time-to-event efficacy variables between the 2 treatment arms:

    - PFS;
    - TTPD.

Safety:

1. To examine the safety and toxicity profile of study treatments, graded according to Common Terminology Criteria for Adverse Events (CTCAE), Version 3.0.

2. To compare hospitalizations, transfusions, and concomitant medication use between the 2 treatment arms.

Quality of Life:

1. To evaluate differences in patient-reported outcomes (PROs), as assessed by the Functional Assessment of Cancer Therapy-Lung/Neurotoxicity (FACT-L/Ntx) instrument (Cella et al. 1995; Calhoun et al. 2003) between the 2 treatment arms.

Pharmacokinetics:

1. To characterize pharmacokinetics (PK) of carboplatin (total and free platinum) and bevacizumab and compare between the 2 treatment arms.

2. To characterize pemetrexed PK in patients randomized to Arm A.

Translational Research:

1. To assess biomarkers relevant to pemetrexed, carboplatin, and bevacizumab.

2. To assess biomarkers relevant to the disease state.

3. To assess the correlation between biomarkers and clinical outcome.

4 INVESTIGATIONAL PLAN

4.1 Overall Study Design and Plan

Study JMHD is a multicenter, randomized, open-label, Phase 3 trial. Eligible patients will be randomized in a 1:1 ratio to one of the two following treatment arms:

Arm A - Pemetrexed, carboplatin, and bevacizumab followed by pemetrexed and bevacizumab (450 patients)

Arm B - Paclitaxel, carboplatin, and bevacizumab followed by bevacizumab (450 patients).

Patients who meet all criteria for enrollment will be randomized to Arm A or Arm B at the baseline visit (Visit 0). Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive voice response system (IVRS) at a central location. Randomization will be stratified by the following factors:

- disease stage (IIIB with pleural effusions versus IV)
- measurable versus non-measurable disease
- ECOG performance status (0 versus 1)
- sex (male versus female).

Treatment will be administered via intravenous (iv) infusion.

Treatment will consist of up to 4 cycles of induction therapy followed by maintenance therapy until disease progression or treatment discontinuation. A maximum of 4 cycles of induction therapy was chosen based on recent reviews and meta-analyses showing this as the optimal duration of initial, platinum-based therapy (as outlined in Section 1.2 of protocol).

4.2 Efficacy and Safety Variables

4.2.1 Primary Efficacy Measure

The primary efficacy measure for this trial is overall survival, defined as the duration, from the date of randomization to the date of death from any cause. If a patient has not died at the time of the data inclusion cutoff date for the analysis, OS will be censored at the last date when the patient was known by the treating physician to still be alive.

4.2.2 Secondary Efficacy Measures

Secondary efficacy measures include overall response rate (RR) and disease control rate (DCR), as well as TTPD and PFS time.

Sites will assess tumor response in patients by using RECIST guidelines (Therasse et al. 2000). Best overall response (BOR), determined from the sequence of cycle responses assessed, will be used for analysis purposes. For BOR of complete response (CR) or partial response (PR), response must be confirmed by a second assessment 28 to 42 days after the first documentation of response. Two objective status determinations of CR before progression are required for a BOR of CR. Two determinations of PR or better before progression, but not qualifying for a CR, are required for a BOR of PR. Best overall response of stable disease (SD) is defined as disease that does not meet the criteria for CR, PR, or PD and has been evaluated at least 1 time, at least 6 weeks after the start of treatment.

The RR will be estimated by dividing the total number of confirmed CRs and PRs by the total number of patients randomized. The DCR will be estimated by dividing the total number of confirmed CRs, PRs, and SDs by the total number of patients randomized.

TTPD: TTPD is defined as the duration from the date of randomization to the first date of objective progression of disease. For patients who receive subsequent systemic anticancer therapy (after discontinuation from the study chemotherapy) prior to objective progression, TTPD will be censored at the date of the last objective progression-free disease assessment prior to starting this subsequent systemic anticancer therapy. For patients not known to have had objective progression of disease as of the data-inclusion cutoff date for a particular analysis, or who have died without objective progression of disease, TTPD will be censored at the date of the last objective progression-free disease assessment prior to that cutoff date or prior to the date of initiation of subsequent systemic anticancer therapy, whichever is earlier. If there is no objective progression-free disease assessment, then TTPD will be censored at randomization date.

PFS: PFS is defined as the duration from the date of randomization to the date of objective progression of disease or the date of death from any cause, whichever is earlier. For patients who receive subsequent systemic anticancer therapy (after discontinuation from the study chemotherapy) prior to objective progression or death, PFS will be censored at the date of the last objective progression-free disease assessment prior to starting the subsequent systemic anticancer therapy. For patients not known to have died as of the data inclusion cutoff date and who do not have objective progressive disease, PFS will be censored at the date of the last objective progression-free disease assessment prior to the cutoff date or prior to the date of initiation of subsequent systemic anticancer therapy, whichever is earlier. If there is no objective progression-free disease assessment, then PFS will be censored at randomization date.

4.2.3 Health Outcome Measures

The Functional Assessment of Cancer Therapy (FACT) questionnaire will be used to capture the patient's assessment of his or her overall quality of life (QOL), patient-reported side effects, and disease and treatment-related symptoms.

This study will use the FACT-General (FACT-G) questionnaire, plus the lung cancer and neurotoxicity subscales. Collectively, the instruments are referred to as the FACT-L/Ntx questionnaire.

The FACT-G is a reliable and valid instrument used to measure QOL in patients with cancer (Cella at al. 1993). The 27-item questionnaire is organized into subscales, each designed to assess a QOL domain:

- physical well-being (PWB) – 7 items
- social/family well-being (SWB) – 7 items
- emotional well-being (EWB) – 6 items
- functional well-being (FWB) – 7 items.

The FACT-Lung (FACT-L) is a reliable and valid instrument consisting of the FACT-G and a 9-item lung cancer subscale (LCS) and is appropriate for both small cell lung cancer and NSCLC (Cella et al. 1995).

The FACT/GOG-Ntx consists of an 11-item neurotoxicity subscale (Ntx) developed in collaboration with the Gynecologic Oncology Group (GOG) for use in patients with neurotoxicity. This instrument possesses strong psychometric properties and is both reliable and valid (Calhoun et al. 2003).

Scores for each subscale (PWB, SWB, EWB, FWB, LCS, Ntx) will be summarized. In addition, the following composite scores will be calculated:

- Total FACT-L: (PWB + SWB + EWB + FWB + LCS)
- Lung Trial Outcome Index (TOI-L): (PWB + FWB + LCS).
- Total FACT/GOG-Ntx: (PWB + SWB + EWB + FWB + NTX)
- Ntx Trial Outcome Index (TOI-Ntx): (PWB + FWB + NTX)

If at least 50% of items within a subscale are completed, each missing item will be imputed by the average of the completed items; if more than 50% of items within a subscale are missing, the score of the subscale will be treated as missing (Cella 1997). The total FACT-L, total FACT/GOG-Ntx, and the TOIs will be scored only if at least 80% of the items that make up these composite scores are completed and at least 50% of the items are completed within each subscale. If these 2 conditions are satisfied, each missing item within each subscale will be imputed by the average of the completed items within the subscale. Otherwise, the total FACT-L, total FACT/GOG-Ntx, and TOI scores will be treated as missing (Fairclough and Cella 1996).

4.2.4 Safety Evaluations

Safety will be assessed by incidence rates of adverse events and serious adverse events, routine serum chemistry and hematology, ECOG performance status, reporting of concomitant medications of interest, and the incidence of hospitalizations and blood transfusions.

4.2.5 Pharmacokinetics/Pharmacodynamics

Pharmacokinetic parameters to be determined by noncompartmental PK analysis based on plasma concentration versus time data for total platinum and free platinum include: $AUC_{0-\infty}$, C_{max}, CL, V_{ss}, and $t_{1/2}$. For bevacizumab, the PK parameters that will be determined include $AUC_{0-\infty}$, C_{max}, and CL. If data do not allow full characterization, then the analysis may be restricted to descriptive evaluation using graphical comparison between Arm A and Arm B.

Pharmacokinetic evaluation of plasma concentration versus time data for pemetrexed may be limited to descriptive evaluation (that is, overlay plot) to verify consistency with reference pemetrexed PK results. In addition, nonlinear mixed effect modeling method (NONMEM®) will be utilized to fit the plasma concentrations of pemetrexed to the best PK model, relying upon PK data and analyses from previous studies. The PK parameters to be estimated include the plasma CL, V, AUC, and $t_{1/2}$.

5 STATISTICAL METHODS

5.1 General Considerations

The missing data of FACT Scores will be imputed according to Section 4.2.3.

Safety and efficacy summaries that are presented by visit will include an End of Study summary, defined as the last available assessment for a given endpoint.

Baseline will be defined as the last assessment prior to receiving the first dose of study medication. Change from baseline will be calculated as the visit value of interest minus the baseline value.

For day to month conversion, a factor of 30.44 will be used.

Data will be summarized according to the visit recorded on the CRF, irrespective of actual elapsed time from the baseline visit or previous visit.

Continuous data will be summarized in terms of the mean, standard deviation, minimum, maximum, quartiles, and number of observations. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, standard deviation, and quartiles will be reported to one more decimal place than the raw data recorded in the database.

Categorical data will be summarized in terms of the number of patients providing data at the relevant time point (n), frequency counts and percentages. Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Unless noted otherwise, percentages will be calculated using n as the denominator.

All confidence intervals will be reported as 95%, two-sided intervals unless stated otherwise. Confidence intervals will be presented to one more decimal place than the raw data.

5.2 Study Patients

5.2.1 Disposition of Patients

A detailed description of patient disposition will be provided. It will include:

- A summary of the number of patients entered into the study and the number and percentage of patients excluded prior to treatment by major reason and overall
- A summary of the number of patients entered into the study, enrolled in the study, and treated, by investigator site and overall
- A summary of the numbers and percentages of patients completing the study and the numbers and percentages of patients discontinuing overall and by reason for discontinuation
- A summary of reasons patients enrolled, but not treated with study drug
- A summary of reasons patients discontinued study treatment
- A summary of all important protocol violations.

By-patient listings of enrollment details, visit dates and withdrawal/study completion details will be provided.

5.2.2 Important Protocol Violations

Important protocol violations are defined in a table which includes these sections:

A summary of the number and percentage of patients with an important protocol violation by treatment group and overall and by type of violation will be provided. And this summary table includes these sections:

- Protocol inclusion/exclusion criteria violations;
- Patients with at least 1 violation related to study discontinuation;
- Incorrect dose modification;

- Treatment violations.

A summary of the number and percentage of patients with at least one violation related to tumor assessment per RECIST guideline by treatment group and by type of violation will be provided.

A by-patient listing of important protocol violations and violations related to tumor assessment per RECIST guideline will be provided.

5.2.3 Analysis Populations

Efficacy analyses will be based on the intent-to-treat (ITT) analysis set. This population is defined as all patients randomized to study treatment. Patients will be grouped according to randomized treatment.

A secondary analysis of the primary efficacy endpoint based upon the Per Protocol Set of patients may be performed if there are significant numbers of patients with important protocol violations (≥ 10% of total patient population). The Per Protocol Set is defined as those patients in the ITT set who are compliant with the study protocol. The precise reasons for excluding patients from the Per Protocol Set will be clearly defined and documented before breaking the blind.

Efficacy analyses may be performed based on safety population if deemed necessary.

Safety analyses will be based on the Safety Population, defined as all enrolled patients receiving at least one dose of any study drug. Patients will be grouped according to actual treatment received.

The pre-baseline summary of distribution of all patients by site and investigator, and summary of patient disposition will be based on the informed consent (IC) population, defined as all patients who provided informed consent for the study.

A by-patient listing of analysis population details will be provided. This listing will include: investigator site, patient identifier, inclusion/exclusion flag for each population and reason for exclusion from each population. All patients screened will appear on this listing.

5.3 Demographic and Other Baseline Characteristics

Patient demographics including age, screening height and weight, and screening Body Surface Area (BSA) will be summarized using descriptive statistics. Summary statistics will include the number of patients (n), mean, standard deviation, minimum and maximum values, and quartiles (i.e., 25^{th}, 50^{th}, and 75^{th} percentiles). Sex and race will be summarized with frequency counts and percentages.

Past and current tobacco use will be summarized by presenting the numbers and percentages of patients who have ever used tobacco products and the numbers and percentages of patients who currently use tobacco products. Average daily consumption

will be categorized as < 5, 5 – 20, > 20. Average daily consumption will be summarized for pipesful of tobacco, cigarettes, cigars, and pinches of smokeless tobacco. Summary statistics will be presented for the number of years patients smoked and for the number of years since quitting for patients who no longer smoke.

Baseline disease characteristics will be summarized by presenting frequency counts and percentages for pathological diagnosis (histological or cytological) and disease stage.

The numbers and percentages of patients reporting prior therapies will be provided overall, and by type of therapy (surgery or radiotherapy). Surgery will be further characterized as palliative or curative. Radiotherapy will be further characterized as neoadjuvant, adjuvant, palliative, or curative.

Historical illnesses that occurred in at least 2% of the patients will be summarized by the numbers and percentages of patients reporting no historical illness at all, at least one such diagnosis, and by Medical Dictionary of Regulatory Activities (MedDRA, Version 14.1) preferred term in alphabetical order.

Pre-existing conditions that occurred in at least 5% of the patients will be summarized by the numbers and percentages of patients reporting no pre-existing conditions at all, at least one such condition, and by Medical Dictionary of Regulatory Activities (MedDRA, Version 14.1) preferred term in alphabetical order.

The numbers and percentages of patients using prior or concomitant medications will be presented. Flagging of concomitant medications as prior, prior and concomitant, or concomitant is discussed in the following section. Medications classified as prior and concomitant will be included in summaries of prior medications and concomitant medications.

By-patient listings of demographic data and baseline patient characteristics will be provided.

5.3.1 Flagging of Concomitant Medications as Prior, Prior and Concomitant or Concomitant

Medication start and stop dates will be compared to the date of first dose of study medication to allow medications to be classified as either Prior, Prior and Concomitant, or Concomitant.

Medications that start and stop prior to the date of first dose of study medication will be classified as Prior medications. If a medication starts before the date of first dose of study medication and stops on or after the date of first dose of study medication then the medication will be classified as Prior and Concomitant. Medications will be classified as Concomitant if they have a start date on or after the date of first dose of study medication.

If medication start and/or stop dates are missing or partial, the dates will be compared as far as possible with the date of first dose of study medication. Medications will be assumed to be concomitant, unless there is clear evidence (through comparison of partial dates) to suggest that the medication started prior to the first dose of study medication. If

there is clear evidence to suggest that the medication started prior to the first dose of study medication, the medication will be assumed to be Prior and Concomitant, unless there is clear evidence to suggest that the medication stopped prior to the first dose of study medication. If there is clear evidence to suggest that the medication stopped prior to the first dose of study medication, the medication will be assumed to be Prior.

5.4 Treatment Compliance

The number of dose omissions, reductions, and delays, and the number of cycles received will be summarized for all treated patients per treatment arm. Summarizations will be performed for the induction period, the maintenance period, and for the combination of the induction and maintenance periods.

A by-patient listing of treatment compliance data will be provided.

5.5 Efficacy Evaluation

5.5.1 Primary Efficacy Evaluation

The primary outcome measure for this trial is OS. All randomized patients, according to the ITT principle, will be included in the analysis of OS. The primary analysis will be the comparison of OS between Arm A and Arm B using a one-sided log-rank test at a significance level of 0.025 (or equivalently, at a two-sided significance level of 0.05).

OS by treatment arm will be summarized by presenting minimum and maximum observed OS, as well as point estimates and 95% CIs for quartiles. Quartiles will be estimated using the product limit method (Kaplan and Meier 1958). Plots of the estimated survival density function over time will be created.

A supportive analysis of the primary outcome will include a multivariate Cox (1972) regression model to estimate treatment hazard ratios (HR) after adjusting for the following potential prognostic variables:

- disease stage (IIIB with pleural effusions versus IV)
- age (<=70 versus > 70 years)
- sex (male versus female)
- ethnic origin (Caucasian versus non-Caucasian)
- basis for pathologic diagnosis (histopathological versus cytological)
- measurable versus nonmeasurable disease
- previously treated brain metastasis (yes versus no)
- smoking status (ever versus never)
- histologic subtype (adenocarcinoma versus large cell versus other)
- ECOG performance status (0 versus 1).

Analyses may exclude any cofactor (listed above) from the Cox models for one or more of the following reasons:

1. If the number of patients representing 1 level of that variable is insufficient

2. If data collected on that variable are insufficiently complete
3. If that cofactor was found to have no prognostic impact ($p > 0.1$).

5.5.2 Secondary Efficacy Evaluation

The secondary efficacy endpoints are RR, DCR, and time-to-event variables PFS and TTPD. All randomized patients, according to the ITT principle, will be included in the analysis of these endpoints.

5.5.2.1 Progression Free Survival

The comparison of PFS between Arm A and Arm B using a one-sided log-rank test will be performed.

K-M plots by treatment arm will be produced, and point estimates and 95% CIs for quartiles will be summarized. In addition, the HR will be estimated using the method described above for OS.

5.5.2.2 Response Rate

The RR will be estimated by dividing the total number of confirmed CRs and PRs by the total number of patients randomized. Exact 95% confidence intervals for the response rate in each treatment arm will be calculated.

5.5.2.3 Time to Progressive Disease

TTPD will be analyzed as described for PFS.

5.5.2.4 Disease Control Rate

The DCR will be estimated by dividing the total number of confirmed CRs, PRs and SDs by the total number of patients randomized. DCR will be analyzed as described for RR.

5.5.3 Subgroup Analyses

5.5.3.1 Time-to-Event Analyses for Patients Who Receive Maintenance Therapy

A subgroup analysis will be performed in the subset of patients who were eligible to receive and subsequently initiated maintenance therapy. As this subset is defined based on post-randomization data, no comparison between treatment arms will be made, but rather a simple descriptive analysis of OS and PFS within each treatment arm will be conducted.

5.5.3.2 Other Efficacy Subgroup Analyses

Subgroup analyses of time-to-event endpoints OS, PFS, and TTPD will be performed for each potential prognostic factor listed in Section 5.5.1 and age categorized by ≤65 versus >65 years old. For each factor, the treatment-by-subgroup interaction will be tested at the 0.10 level of significance using a Cox regression model with terms for treatment, the factor, and treatment-by-factor interaction. The log-rank test will be performed, and Kaplan-Meier estimates and HR estimates from a univariate Cox regression model will

be calculated in order to estimate and compare treatment effect in each subgroup category, for each factor found to have a significant treatment-by-subgroup interaction.

Subgroup analyses of RR and DCR will also be performed for each potential prognostic factor listed in Section 5.5.1 and age categorized by ≤65 versus >65 years old. For each factor, the treatment-by-subgroup interaction will be tested at the 0.10 level of significance using a logistic regression model with terms for treatment, the factor, and treatment by the factor interaction. Then, Fisher exact test will be performed to compare treatment difference in each subgroup category, for each factor found to have a significant interaction term.

5.5.3.3 Other Safety Subgroup Analyses

If deemed appropriate, subgroup analyses of selected AEs may be performed for each potential prognostic factor listed in Section 5.5.1 and age categorized by ≤65 versus >65 years old. For each factor, the treatment-by subgroup interaction will be tested at the 0.10 level of significance using logistic regression model containing treatment, the factor, and treatment by the factor interaction in the model. Then, Fisher exact test will be performed to compare treatment difference in each subgroup category.

5.5.4 Sensitivity Analyses for PFS, TTPD, and PFS without Grade 4 Toxicity (G4PFS)

PFS: Duration is measured from the date of randomization to the date of objective progression of disease or the date of death from any cause, whichever is earlier.

There will be two sets of sensitivity analyses for PFS based on two different censoring rules.

Censor Rule #1:
For patients not known to have died as of the data inclusion cutoff date and who do not have objective progressive disease, PFS will be censored at the date of the last objective progression-free disease assessment prior to the cutoff date. Basically the "subsequent systemic anticancer therapy" criterion is removed from the primary analysis. If there is no objective progression-free disease assessment, then PFS will be censored at randomization date.

Censor Rule #2:
1. For patients without baseline tumor assessment, PFS will be censored at randomization date.

2. For patients who have two consecutive missing tumor assessments prior to objective progression or death from any cause, PFS will be censored at the last assessment date prior to the missing assessment.

3. For patients who receive subsequent systemic anticancer therapy (after discontinuation from the study chemotherapy) prior to objective progression or death, PFS will be censored at the date of the last objective progression-free disease assessment prior to starting the subsequent systemic anticancer therapy.

4. For patients not known to have died as of the data inclusion cutoff date and who do not have objective progressive disease, PFS will be censored at the date of the last objective progression-free disease assessment prior to the cutoff date or prior to the date of initiation of subsequent systemic anticancer therapy, whichever is earlier. If there is no objective progression-free disease assessment, then PFS will be censored at randomization date.

TTPD: Duration is measured from the date of randomization to the first date of objective progression of disease.
There will be two sets of sensitivity analyses for TTPD based on two different censor rules.

Censor Rule #1:
For patients not known to have had objective progression of disease as of the data-inclusion cutoff date for a particular analysis, or who has died without objective progression of disease, TTPD will be censored at the date of the last objective progression-free disease assessment prior to the cutoff date. Basically the "subsequent systemic anticancer therapy" criterion is removed from the primary analysis. If there is no objective progression-free disease assessment, then TTPD will be censored at randomization date.

Censor Rule #2:
1. For patients without baseline tumor assessment, TTPD will be censored at randomization date.

2. For patients who have two consecutive missing tumor assessments prior to objective progression or death due to cause of disease progression, TTPD will be censored at the last assessment date prior to the missing assessment.

3. For patients who receive subsequent systemic anticancer therapy (after discontinuation from the study chemotherapy) prior to objective progression, TTPD will be censored at the date of the last objective progression-free disease assessment prior to starting the subsequent systemic anticancer therapy.

4. For patients not known to have had objective progression of disease as of the data-inclusion cutoff date for a particular analysis, or who have died without objective progression of disease, TTPD will be censored at the date of the last objective progression-free disease assessment prior to that cutoff date or prior to the date of initiation of subsequent systemic anticancer therapy, whichever is earlier. If there is no objective progression-free disease assessment, then TTPD will be censored at randomization date.

G4PFS:
All randomized patients, according to the ITT principle, will be included in the analysis of G4PFS, which will be measured from the date of randomization to the earliest

occurrence date of one of the following three events: CTC grade 4 adverse events, disease progression or death from any cause. For patients who receive subsequent systemic anticancer therapy prior to any of the events, G4PFS will be censored at the date of the last objective progression-free disease assessment prior to starting the subsequent systemic anticancer therapy. For patients not known to have any of the events as of the data inclusion cutoff date, G4PFS will be censored at the date of the last objective progression-free disease assessment prior to the cutoff date or prior to the date of initiation of subsequent systemic anticancer therapy, whichever is earlier.

5.6 Health Outcome Analyses

Each of the subscales and composite scores defined in Section 4.2.3 will be summarized using descriptive statistics at each protocol specified visit. Change from baseline will be summarized as well. Treatment group differences in change from baseline will be compared using an analysis of covariance (ANCOVA) model with terms for treatment and baseline value. Least squares (LS) means for the differences between treatment groups in change from baseline will be presented, along with associated standard errors and 95% CIs.

Post-baseline values will be categorized as improved, worsened, or no change using the criteria for clinically meaningful change in Table JMHD.6 in the protocol. For scales for which a clinically meaningful change is not defined, categorization will be determined by an increase or decrease from baseline of at least ½ standard deviation, where the standard deviation is estimated from the baseline scores of all ITT population per treatment. If a change from baseline value should fail to meet the criterion for clinically meaningful change but exceed (in absolute value) ½ the standard deviation estimate, it will be categorized as having worsened or improved, accordingly. The numbers and percentages of patients categorized as improved, worsened, or unchanged will be presented at each post-baseline visit at which the assessment was performed.

Other exploratory analyses may be performed, including longitudinal modeling (for example, repeated measures models and the impact of covariates), subgroup analysis (for example, age ≤ 70 versus > 70, and subgroups based on ECOG performance status, tumor response status, PFS, and OS), association between clinical outcome and health outcome measurements and factor analysis (for example, to explore which FACT items load on neuropathy, fatigue, and alopecia).

Raw responses to the FACT-L/Ntx instruments will be included in patient listings. Derived scores will be included as well.

5.7 Safety Evaluation

All safety summaries and analyses will be based upon the Safety Population as defined in Section5.2.3.

5.7.1 Extent of Exposure

The numbers and percentages of patients completing one cycle, two cycles,…, 10 cycles, > 10 cycles will be presented. In addition, summary statistics (mean, median, minimum and maximum) will be presented for the numbers of cycles.

For pemetrexed actual dose in mg/m^2 will be calculated as actual dose (mg) divided by BSA (m^2). Actual dose for paclitaxel will be calculated similarly. Actual dose for bevacizumab in mg/kg will be calculated as actual dose (mg) divided by weight (kg). Planned dose and actual dose of carboplatin (mg) will be based on the values recorded on the CRF. Mean planned dose and mean actual dose will be calculated per patient per week as the sum of doses divided by 3 × number of cycles in a given treatment period. Summary statistics will be presented by treatment period for mean planned dose, mean actual dose and percent dose intensity, defined as 100 × mean actual dose/mean planned dose.

The number and percentage of patients with any dose adjustment will be presented. Dose adjustments will be further characterized by presenting the numbers and percentages of patients with dose decreases and dose omissions. Reasons for dose adjustments will be summarized as well.

Details of study drug administration will be included in patient listings.

5.7.2 Adverse Events

MedDRA (Version 14.1) will be used to group adverse events by System Organ Class (SOC) and Preferred Term (PT).

Treatment-emergent adverse events (TEAEs) will be tabulated, i.e., those events that were not present at baseline. An AE that was present at baseline will be considered to be treatment-emergent if it (1) worsened in CTCAE grade following the start of treatment, (2) is reported to be possibly related to study drug or procedure subsequent to start of treatment or (3) was non-serious at baseline and becomes serious after the start of treatment. Any adverse events starting after the 30 day post-discontinuation follow-up period will be included in patient listings but will not be tabulated.

The onset date of an adverse event will be compared to the date of first dose of study medication to determine if the adverse event is treatment emergent or not. Adverse events with an onset date on or after the date of first dose of study medication will be classified as treatment-emergent.

If the adverse event onset date is missing or partial, the date will be compared as far as possible with the date of first dose of study medication. Adverse events will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the adverse event started prior to the first dose of study medication.

An overall summary of adverse events will be provided. Frequency counts and percentages will be used to summarize the following:

- Patients with at least one TEAE
- Patients with at least one CTCAE Grade 3 or 4 TEAE
- Patients with at least one serious adverse event (SAE)
- Patients who discontinued due to non-serious AE
- Patients who discontinued due to SAE
- Patients who died on therapy
- Patients who died within 30 days of last dose of study drug

The above summaries will be provided for all events, regardless of study drug causality, and repeated for events deemed by the investigator to be possibly related to study medication. Incidence rates of these events will be compared between treatment arms using Fisher's exact test.

Treatment emergent AEs will be summarized by SOC and by decreasing frequency of PT within each SOC. TEAEs will be further summarized by PT and maximum CTCAE grade. In addition, laboratory and non-laboratory events will be summarized by CTCAE term and maximum CTCAE grade.

A by-patient listing of all adverse events (including non-treatment-emergent events) will be provided. This listing will include investigator site, patient identifier, age, sex, race, adverse event (SOC, PT, reported term), event start date, event end date, duration, CTCAE grade, relationship to study drug/procedure, seriousness, and outcome.

For each patient and each adverse event, the worst severity recorded will be attributed and used in the by-severity summaries. Similarly, the worst causality (most related to treatment) will be attributed and used in the by-causality summaries. If severity or causality is missing, the worst case will be assumed.

5.7.3 Deaths, Serious Adverse Events, and Other Significant Adverse Events

Reasons for deaths will be summarized separately for on-therapy and within 30 days of last dose of study drug. A by-patient listing of mortality status will be provided.

Serious adverse events will be summarized by PT. SAEs will be summarized overall and by cycle. A listing of serious adverse events will be produced following the format described for adverse events in Section 5.7.2.

5.7.4 Clinical Laboratory Evaluation

Clinical laboratory assessments performed by a central lab, i.e., routine chemistry, will be summarized. Local lab results, i.e., hematology, will be included in patient listings but will not be tabulated.

A by-patient listing of all laboratory data will be provided, with abnormal values highlighted, and will include investigator site, patient identifier, age, sex, race, weight and visit. Laboratory reference ranges will also be listed.

5.7.5 Other Observations Related to Safety

The numbers and percentages of patients with any hospitalizations during the study treatment period or during the 30-day post-discontinuation follow-up period will be summarized by treatment group. Hospitalization incidence rates will be compared between the treatment groups using Fisher's exact test. In addition, total number of days in hospital will be summarized. Differences between treatment groups in hospitalization days will be compared using the Wilcoxon rank sum test.

The numbers and percentages of patients with any blood transfusions during the study treatment period or during the 30-day post-discontinuation follow-up period will be summarized by treatment group. Transfusions will be further characterized by transfused blood product, i.e., packed red blood cells, platelets, fresh frozen plasma, or whole blood. The proportions of patients having blood transfusions will be compared between the treatment groups using Fisher's exact test.

Use of concomitant medications of interest (for example, erythropoiesis stimulating agents, growth factors, antibiotics, anti-seizure agents, antidepressants, and analgesics) during the study treatment period or during the 30-day post-discontinuation follow-up period will be summarized by presenting the numbers and percentages of patients reporting any such use. The proportions of patients reporting use of medications of interest will be compared between the treatment groups using Fisher's exact test.

The numbers and percentages of patients reporting post-discontinuation therapies will be provided overall, and by type of therapy (surgery, radiotherapy, or systemic therapy). Intent of surgery will be characterized as palliative or curative. Reason for radiotherapy will be characterized as neoadjuvant, adjuvant, palliative, locally advanced or metastatic. Systemic therapy will be further categorized by WHO drug terms.

Details of hospitalizations, transfusions, use of concomitant medication of interest, and post-discontinuation therapies will be included in patient listings.

5.7.6 Sensitivity Analyses for Adverse Events

Sensitivity analyses will be conducted to evaluate the impact of any missing CTCAE grades on the overall assessment of safety for the study. If any CTCAE grade values are missing at baseline, the missing grade will be assigned as 1. If any CTCAE grades are missing at a post-baseline visit, the following three imputation scenarios will be used:

1) Impute missing grade as 4 in Arm A only, no imputation on Arm B will be performed;
2) Impute missing grade as 4 in Arm A and B;

3) Impute missing grade (in Arm A only) by the worst grade among those patients with the same CTCAE term (in post baseline) that do not have a missing grade (in Arm A only). If the worst grade is 5, impute missing grade as 4.

Additional sensitivity analyses such as excluding patients with missing CTCAE grade from the analysis set may be conducted. The adverse events with missing CTCAE grade will be included in the patient listing.

5.8 Pharmacokinetic/Pharmacodynamic Analyses

5.8.1 Pharmacokinetic Parameter Estimation

Pharmacokinetic parameters will be computed using noncompartmental methods of analysis using WinNonlin® on a computer that meets or exceeds the minimum system requirements for this program. The version of any software used for the analysis will be documented and the program will meet the Lilly requirements of software validation. It is possible that other validated equivalent PK software programs may be utilized if appropriate, warranted, and approved by Lilly Global Pharmacokinetic management.

Pharmacokinetic parameters to be determined by noncompartmental PK analysis based on plasma concentration versus time data for total platinum and free platinum include: $AUC_{0-\infty}$, Cmax, CL, Vss, and $t_{1/2}$. For bevacizumab, the PK parameters that will be determined include $AUC_{0-\infty}$, Cmax, and CL. If data do not allow full characterization, then the analysis may be restricted to descriptive evaluation using graphical comparison between Arm A and Arm B.

For pemetrexed, the PK evaluation of plasma concentration versus time data may be limited to descriptive evaluation (that is, an overlay plot) to verify consistency with reference Pemetrexed concentrations. The nonlinear mixed effect modeling method will be implemented using the NONMEM® software to further evaluate the plasma concentrations of pemetrexed.

Further details regarding the PK analysis are included in the Lilly PK analysis plan.

5.8.2 Pharmacokinetic Statistical Inference

Free platinum, total platinum, and bevacizumab dose-normalized $AUC_{0-\infty}$, (or partial AUC based on plasma concentration time-profile) and dose-normalized Cmax will be log-transformed and analyzed using a linear mixed-effects model allowing for the fixed effect of treatment (such as, carboplatin with [Arm A] or without pemetrexed [Arm B], or bevacizumab with [Arm A] or without pemetrexed [Arm B]) and the random effect of patient.

Least squares means will be calculated for free platinum, total platinum and bevacizumab AUC in Arms A and B, as well as the mean differences between both treatments (with or without pemetrexed). The residual variance from the analysis of variance (ANOVA) will

be used to calculate the 90% confidence intervals for the mean differences between treatments. These values will be back transformed to give point estimates and 90% confidence interval for the ratios of the treatments. For these calculations, Arm B (absence of pemetrexed) will be taken as the reference. If the 90% confidence interval for the ratio of AUCs is contained between 0.7 and 1.43, then it could be concluded that no statistically significant drug-drug interaction exists between carboplatin and pemetrexed or between bevacizumab and pemetrexed.

Variance components for the inter-patient (between) and intra-patient (within) variability in Cmax and AUC will be obtained through the linear mixed-effects model. These variances will be expressed as coefficients of variation. The procedure MIXED in the statistical package SAS® will be utilized in running the ANOVA models and in estimating the variance components

5.9 Translational Research

The distributions of biomarkers with continuous measures, such as gene or protein expression, will be described for the total patient population and by treatment arm. Summary statistics will include means, medians, corresponding standard deviation, quartiles, and ranges. Biomarkers with discrete measures, such as genotype locus or IHC-staining assessed protein expression, will be summarized in frequency tables for the total population and by treatment arm..

Associations between clinical endpoints and markers will be evaluated using data from patients who have a reported value for the marker measure of interest. The detailed description of translational research analyses are captured in a separate biomarker statistical analysis plan.

5.10 Patient Profiles

No patient profiles will be provided.

5.11 Determination of Sample Size

The study will enroll approximately 900 patients (450 per arm), with the final analysis of OS to be performed after 676 deaths have occurred. Assuming an HR of 0.80, this sample size yields at least 80% statistical power to detect superiority of the pemetrexed-combination arm over the paclitaxel-combination arm with the use of a 1-sided log-rank test and a type I error of 0.025 (Freedman 1982). If the true median OS for the paclitaxel-combination arm (Arm B) is 12.3 months (Sandler et al. 2006), then the HR of 0.8 amounts to an approximately 3-month improvement in median OS for the pemetrexed-combination arm (Arm A), under assumption of exponential survival distribution. Assuming approximately 5% screening failure, the study will enter approximately 950 patients.

6 INTERIM ANALYSIS

Four planned interim analyses will be conducted for this study.

1. The first interim analysis will be conducted for safety and futility after 250 patients have been enrolled to the study. The data cutoff date for inclusion in this interim analysis is the date when the 250th patient has completed four cycles of treatment or has been followed up for three months since enrollment if the patient is discontinued before four cycles of treatment. All visits completed prior to this data cutoff date will be validated for inclusion in the interim database.
2. The second interim analysis will be conducted for safety when 500 patients have been enrolled to the study. The data cutoff date for inclusion in this interim analysis is the date when the 500th patient has completed one cycle of treatment. All visits completed prior to this data cutoff date will be validated for inclusion in the interim database.
3. The third interim analysis will be conducted for safety when 750 patients have been enrolled to the study. The data cutoff date for inclusion in this interim analysis is the date when the 750th patient has completed one cycle of treatment. All visits completed prior to this data cutoff date will be validated for inclusion in the interim database.
4. The fourth and final interim analysis will be conducted for safety after all planned 900 patients have been enrolled to the study. The data cutoff date for inclusion in this interim analysis is the date when the last patient has completed one cycle of treatment. All visits completed prior to this data cutoff date will be validated for inclusion in the interim database.

During each interim lock and analysis period, patient enrolment will continue but will not be included in that interim analysis.

An external Data Monitoring Committee (DMC) will be formed and will be responsible for evaluating interim results. Only the DMC is authorized to review completely unblinded interim efficacy and safety analyses. The DMC will be responsible for making recommendations as to whether continuing study treatment and enrollment is scientifically and ethically appropriate. A list of the planned interim analyses will be included in the DMC charter. A separate interim statistical analysis plan (iSAP) will be created after DMC charter approval to contain mocked tables, figures, and listings for the planned interim analyses as well as details concerning treatment unblinding. This iSAP will be a living document containing planned as well as any additional interim analyses the DMC may request prior to interim lock.

Stopping Guidance for Safety

There will be no pre-specified rules for stopping the trial due to safety concerns. DMC members will review unblinded safety data at each interim analysis to determine whether

there are sufficient safety concerns to justify the termination of study treatment and/or enrollment.

Stopping Guidance for Futility at the First Interim

Futility for the first interim analysis will be determined in terms of PFS because the number of deaths is likely to be too low to yield a meaningful analysis on OS. The futility rule will be based on conditional power, which is the probability of observing significant results at the final analysis under alternative hypothesis assumption, given data at the interim. Because the alternative hypothesis at final analysis is based on the HR for OS, a corresponding alternative hypothesis in terms of HR for PFS is needed in order to calculate the conditional power in terms of PFS. It is believed that if the pemetrexed-combination arm (Arm A) can reduce the risk of death by 20% (that is, the alternative HR 0.8 for OS), then it should reduce risk of progression by at least 10%. Therefore, the conditional power will be calculated under the alternative hypothesis of HR 0.9 for PFS.

As guidance, if the conditional power is less than 12% (for illustrative purpose, this corresponds to the PFS HR > 1.21 under a Cox PH model), then the trial would be considered futile. This futility analysis will not affect overall alpha level and hence no alpha adjustment is needed for this interim analysis.

The futility threshold value of 12% was determined by a simulation study. In the simulation, the exponential PFS data with median PFS time of 6.2 months for the paclitaxel-combination arm (Arm B; Sandler et al. 2006) and median PFS time of 6.9 months for the pemetrexed-combination arm (Arm A) were generated according to the alternative hypothesis of HR 0.9. If 250 patients can be enrolled in the first 11 months and all 900 patients can be enrolled in 24 months, and if the interim analysis will be performed three months after the enrollment of 250 patients, then the conditional power falls below 12% in 5% of 1000 simulations. Similar simulation results were obtained if enrollment of 900 patients takes up to 30 months.

Because the stopping guidance for this interim analysis was based on a simulation, which was based on a number of assumptions (such as exponential PFS, accrual time, and alternative hypothesis of HR 0.9 for PFS) that may well be violated by the actual trial enrollment and the observed data, the stopping guidance should be viewed as only guidance, not the absolute rules, and they may be updated in a separate iSAP prior to an interim data lock if deemed necessary.

This is an open-label study. Each patient will be aware of his or her own assigned treatment group. At each investigative site, all staff involved in treating and caring for study patients will have full knowledge of treatment assignments for the patients under their care.

For the creation of analysis data sets (ADS) supporting the interim analyses, treatment assignment and other parameters that can disclose treatment assignment will be scrambled within the extract data sets (EDS) until the final database lock. Programming

of ADS and interim tables, listings and figures (TLFs) will be based upon the scrambled EDS. Sponsor review of interim ADS and TLFs also will be based upon the scrambled EDS. Therefore, the sponsor and all PAREXEL Biostatistics and Clinical Programming staff will remain blinded to treatment group assignments for the aggregate database until the database lock for the final analysis.

The following data elements will be scrambled or masked in order not to disclose treatment assignment:

- Subject ID will be scrambled in every EDS. In particular, reordering subject ID in the treatment assignment EDS will result in an arbitrary association between subject and treatment, as illustrated in the following example.

Before

Subject ID	Treatment Assignment
1	Arm A
2	Arm B
3	Arm B
4	Arm A

After

Subject ID	Treatment Assignment
4	Arm A
2	Arm B
1	Arm B
3	Arm A

- Study Drug Therapy – Paclitaxel: Variables Dose Adjustment Type and Reason for Dose Delay will be scrambled. Variables Adverse Event ID, Dose Reduced Event ID, and Dose Discontinued, AE Event ID will be set to missing. Variable Dose Administered will be assigned an arbitrary value of 750 mg.
- Study Drug Therapy – Pemetrexed: Same as for Paclitaxel.
- Subject Status - Arm A: Variable Subject Status at the End of This Cycle will be scrambled
- Subject Status: Arm B: Same as for Arm A.

7 REFERENCES

Calhoun EA, Welshman EE, Chang CH, Lurain JR, Fishman DA, Hunt TL, Cella D. 2003. Psychometric evaluation of the Functional Assessment of Cancer Therapy/ Gynecologic Oncology Group-Neurotoxicity (Fact/GOG-Ntx) questionnaire for patients receiving systemic chemotherapy. Int J Gynecol Cancer 13(6):741-748.

Cella D. 1997. Manual of the Functional Assessment of Chronic Illness Therapy (FACIT), Version 4. Center on Outcomes, Research and Education (CORE). Evanston Northwestern Healthcare and Northwestern University.

Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, Wolf MK, Johnson DH. 2002a. What is a clinically meaningful change on the Functional

Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. J Clin Epidemiol 55(3):285-295.

Cella DF, Bonomi AE, Lloyd SR, Tulsky DS, Kaplan E, Bonomi P. 1995. Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. Lung Cancer 12(3):199-220.

Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, Silberman M, Yellen SB, Winicour P, Brannon J, Eckberg K, Lloyd S, Purl S, Blendowski C, Goodman M, Barnicle M, Stewart I, McHale M, Bonomi P, Kaplan E, Taylor S IV, Thomas CR Jr, Harris J. 1993. The Functional Assessment of Cancer Therapy (FACT) scale: development and validation of the general measure. J Clin Oncol 11(3):570-579.

Cox DR. 1972. Regression models and life-tables. J Royal Stat Soc Ser B 74(2):187-220.

Dmitrienko A, Tamhane AC. 2007. Gatekeeping Procedures with Clinical Trial Applications. Pharmaceutical Statistics 6: 171-180.

Fairclough DL, Cella DF. 1996. Functional Assessment of Cancer Therapy (FACT-G): non-response to individual questions. Qual Life Res 5(3):321-329.

Freedman LS. 1982. Tables of the number of patients required in clinical trials using the logrank test. Stat Med 1(2):121-129.

Jennison C, Turnbull BW. 1999. Group Sequential Methods with Applications to Clinical Trials. Boca Raton (FL): Chapman and Hall/CRC.

Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457-481.

Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, Lilenbaum R, Johnson DH. 2006. Paclitaxel-carboplatin alone or with bevacizumab for non-small cell lung cancer. N Engl J Med 355(24):2542-2550.

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG. 2000. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 92(3):205-216.

8 REPORTING OUTPUT

All outputs will be produced using SAS® version 9.1.3 or a later version (if available at PAREXEL). The REPORT procedure will be used to produce all tables and listings whenever possible. The GPLOT procedure will be used to produce all figures whenever possible. All statistical appendices (supportive SAS output) will be output directly from the appropriate SAS procedure.

Post-text tables, listings and statistical appendices will be produced as RTF files using ODS and Courier New font size 9. Figures will be produced as RTF files using ODS and font=simplex. For all outputs, the page numbering will be applied to ensure that when the RTF files are combined, the page numbering remains fixed.

PK appendices, tables, figures, and listings may be generated from validated software, such as S-Plus, SigmaPlot, or WinNonlin, with file types and formats consistent with local Lilly PK standards.