CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

205832Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 205,832/0000
Drug Name:	Ofev® (nintedanib) 150 mg capsules
Indication(s):	Treatment of Idiopathic Pulmonary Fibrosis (IPF)
Applicant:	Boehringer Ingelheim Pharmaceuticals, Inc.
Date(s):	Submitted: May 2, 2014
	PDUFA: January 2, 2015
Review Priority:	Priority
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Yongman Kim, Ph.D.
Concurring Reviewers:	David Petullo, M.S.
	Thomas Permutt, Ph.D.
Medical Division:	Division of Pulmonary, Allergy, and Rheumatology Products
Clinical Team:	Miya Paterniti, M.D.
	Banu Karimi-Shah, M.D.
Project Manager:	Jessica Lee, Pharm. D.

Keywords: NDA, clinical studies, missing data, random coefficient regression model

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Boehringer Ingelheim Pharmaceuticals, Inc. has proposed Ofev® (nintedanib) for the treatment of idiopathic pulmonary fibrosis (IPF) [redacted](b) (4) IPF occurs primarily in older adults, is limited to the lungs, and is a fatal disease. There are currently no drugs approved for the treatment of IPF in the United States. The applicant submitted the results from two phase 3 clinical trials and one phase 2 trial to support the efficacy of nintedanib for the treatment of IPF. The applicant claims that the results from these trials provide substantial evidence of efficacy by slowing the rate of decline in Forced Vital Capacity (FVC).

Based on my review of the data from the two phase 3 studies, 1199.32 and 1199.34, and the one phase 2 study, 1199.30, there is sufficient evidence to support the efficacy of nintedanib 150 mg in treating patients with IPF. For the three studies reviewed, the analysis of the predefined primary efficacy endpoint, annual rate of decline in lung function, was statistically significant. This evidence was further supported by the analyses of secondary endpoints in studies 1199.34 and 1199.30. In these studies, 1199.34 and 1199.30, the key secondary endpoints, Saint George's Respiratory Questionnaire (SGRQ) total score and time to first acute IPF exacerbation, were statistically significantly in favor of nintedanib. As additional evidence of efficacy, the results from these three studies were integrated and examined for a difference in all-cause mortality. From a clinical viewpoint, mortality is the primary endpoint of interest in patients with IPF. Results indicated a numerical trend favoring nintedanib, although not statistically significant. Therefore, from a statistical perspective, the overall package provided substantial evidence of nintedanib's efficacy benefit.

2 INTRODUCTION

2.1 Overview

This application was submitted on May 2, 2014 in support of nintedanib 150 mg bid for the treatment of patients with idiopathic pulmonary fibrosis. Nintedanib is a small molecule tyrosine kinase inhibitor including the receptors platelet-derived growth factor receptor α and β, fibroblast growth factor receptor 1-3, and vascular endothelial growth factor receptor. In this NDA, the applicant is seeking marketing approval of nintedanib for treatment of idiopathic pulmonary fibrosis. IPF is a specific form of chronic, progressive fibrosing interstitial pneumonia of unknown cause that occurs primarily in older adults and is a fatal disease.

The submission included the results from two phase 3, randomized, double-blind, placebo-controlled studies, 1199.32 and 1199.34, that were identical in design and a phase 2, randomized, double-blind, placebo-controlled dose-ranging study, 1199.30. The objective of phase 3 studies was to evaluate the efficacy and safety of nintedanib 150 mg bid treatment compared with placebo in patients with IPF. The objective of phase 2 study was to evaluate the efficacy and safety of 4 dose strategies of nintedanib treatment compared with placebo in patients with IPF. In each study, patients were to receive randomized, double-blind study treatment for 52 weeks. The primary efficacy outcome variable was the rate of decline in FVC from Baseline to Week 52.

History of Drug Development and Regulatory Interactions

The nintedanib clinical development program was first introduced to the Division of Pulmonary, Allergy, and Rheumatology Products in 2007 under IND 74,683. Communication with the applicant regarding their development plan is documented under this IND. Pertinent parts of the statistical portion of those communications are summarized herein.

In December 2007, the applicant had a Pre-IND meeting with the Division, where input was received regarding the proposed phase 3 program. The Division provided the following statistical comments on the proposed analysis plan:

> *A. The analysis plan for the yearly rate of decline in FVC is unclear. Portions of the briefing package indicate that the response variable of interest is the within subject slope associated with a linear function of FVC by time from first treatment intake to the last much like that used for estimation of growth velocity in a growth study.*
> *(…)*
> *B. If this study may be relied upon as a confirmatory trial for regulatory purposes, the protocol should address the following.*
>
> > *(a) Adjustment for multiple comparisons (i.e., multiple doses) is needed.*
> > *(b) Analyses incorporating data collected past one year in this study may be biased in that only a subset of subjects will enter that portion of the study (i.e., only those who well tolerate and benefit from study medication).*
> > *(c) Methods for imputation of missing data should be clearly justified and sensitivity analyses to assess the appropriateness of the assumptions made in this context should be proposed. The protocol indicates that FVC is expected to decline over the course of the study brining into question the appropriateness of the last-observation-carried-forward approach for even a subset of subjects.*
> > *(…)*

In December 2010, the applicant had an EOP2 meeting with the Division, where input was received regarding the proposed phase 3 program. The Division provided the following statistical comments on the proposed protocols and analysis plan:

- *The Division considered the primary and the key secondary endpoints to appear as reasonable.*
- *The Division noted that mortality was an important clinical endpoint to be evaluated regardless of the results of the analyses of primary and secondary efficacy endpoints.*
- *The Division asked to pre-specify the variance-covariance structure in the primary statistical model.*
- *The Division commented on the MAR assumption as a potential review issue and recommended sensitivity analyses.*
- *The Division noted that study powering was based upon a difference of 100 mL and commented that the assessment of efficacy will evaluate the statistical significance and clinical meaningfulness of the treatment effect.*

In June 2011, Orphan Drug Designation to nintedanib for the treatment of patients with IPF was granted. As IPF is a life-threatening disease and given the serious unmet medical need in the US, Fast Track designation to nintedanib in IPF was granted in May 2013.

In August 2013, the applicant received a Type C meeting written response from the Division, where input was received regarding the proposed phase 3 program. The Division provided the following statistical comments on the proposed analysis plan:

- *The Division recommended an event driven analysis with mortality as the event of interest.*
- *The Division commented on the proposed sensitivity analyses using Pattern Mixture Model for missing data.*

2.1.1 Specific Studies Reviewed

The focus of this review is on the efficacy data from two phase 3 efficacy studies, 1199.32 and 1199.34, and one phase 2 study 1199.30. The design of the three studies, which is also referenced in the label, is described in Table 1.

Table 1. Clinical Trials Reviewed

BI Trial No.	Phase	Design	Treatment Arms	Number of Patients	Dates
1199.32	3	52-week, randomized, double-blind, parallel-group, placebo-controlled	Nintedanib 150 mg bid Placebo	309 206	05/2011-10/2013
1199.34	3	same as 1199.32	Nintedanib 150 mg bid Placebo	331 220	05/2011-10/2013
1199.30	2	52-week, dose finding, proof-of concept, randomized, double-blind	Nintedanib 50 mg qd Nintedanib 50 mg bid Nintedanib 100 mg bid Nintedanib 150 mg bid Placebo	87 86 86 86 87	09/2007-06/2010

Source: Reviewer

2.2 Data Sources

NDA 205-832 can be found in the electronic document room (EDR) of the Center for Drug Evaluation and Research. The study report including protocols, statistical analysis plan, and all referenced literature can be found in the EDR. The program codes used in statistical analyses and the electronic data sets with raw and derived variables and data definitions were provided in the EDR using the following path:

\\cdsesub1\evsprod\NDA205832\0000\m5\datasets
\\cdsesub1\evsprod\NDA205832\0018\m5\datasets

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

In general, the submitted efficacy data were acceptable in terms of quality and integrity. I was able to derive the primary and secondary efficacy endpoints for the studies reviewed. No noticeable deviations between the case report forms and analysis datasets relevant to primary and secondary endpoints were identified. The statistical analyses of my derived endpoints were consistent with the applicant's analyses.

Based on the information provided in this submission, each study seemed to be conducted properly and was consistent with the history of regulatory interactions, protocol revisions/amendments,

study report, and study datasets. The Office of Scientific Investigations had not finalized their inspection of this application at time of my review.

3.2 Evaluation of Efficacy

Studies 1199.32 and 1199.34 were of identical design and will be discussed in Section 3.2.1. The phase 2 study, 1199.30, will be discussed separately in Section 3.2.2. For simplicity, nintedanib 150 mg bid will be denoted by nintedanib.

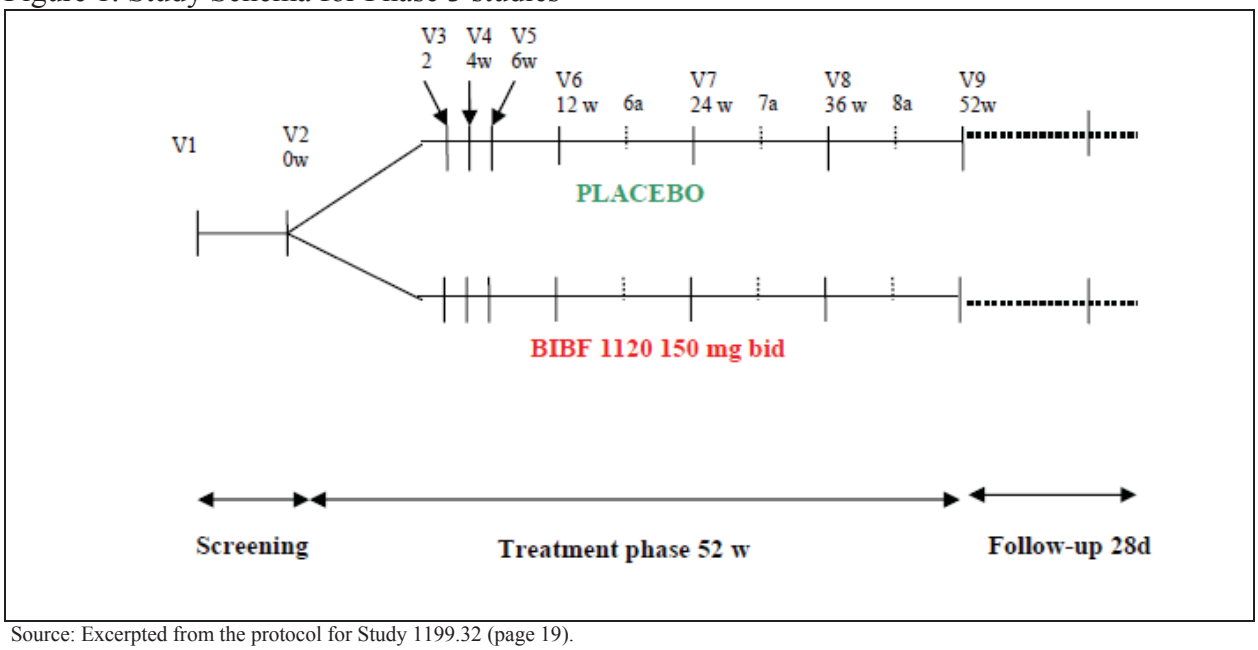
3.2.1 Studies 1199.32 and 1199.34

The applicant conducted two phase 3, randomized, double- blind, placebo-controlled international studies, Study 1199.32 and Study 1199.34. In all study design features (eligibility criteria, study schedule, primary efficacy outcome variable and analysis, secondary and exploratory efficacy outcome measures and analyses, and all safety outcome measures and analyses), the studies were identical. The objective of the studies was to evaluate the efficacy and safety of nintedanib 150 mg bid compared with placebo in patients with IPF. In each study, patients were to receive randomized study treatment in a double blind manner for 52 weeks.

Study Design and Endpoints

These studies consisted of a screening period, a double-blind treatment period, and a final follow-up visit, Figure 1. After a screening visit, if the patient complies with all inclusion and exclusion criteria, randomization was performed by phone or Internet. Patients then entered a 52 weeks treatment phase. Nine visits (visit 2 to 9 + follow up) were planned within one year of treatment, and intermediate lab tests were planned when the interval between two visits increases.

Figure 1. Study Schema for Phase 3 studies



Source: Excerpted from the protocol for Study 1199.32 (page 19).

The diagnosis of IPF were confirmed by central review of high resolution computerized tomography (HRCT) and, if available, surgical lung biopsy. Patients were required to be aged≥40, diagnosed with IPF within last 5 years, and to have %Predicted forced vital capacity (FVC)

≥50% and %Predicted carbon monoxide diffusing capacity (DL_{CO}) 30% to 79%.

Spirometry measurements, including FVC and FEV_1 were to be assessed at Screening, Day 1 (before randomization), at Week 2, 4, 6, 12, 24, 36, 52 (or End of Treatment), and the Final Follow-up visit. At each visit, three FVC values were collected before and after bronchodilator, respectively, until maximum acceptable FVC value was chosen.

Statistical Methodologies

The primary analysis population was the Treated Set (TS) defined as all randomized patients who received any amount of study treatment. Of note, there were two placebo patients who were not treated with study drug.

The primary efficacy outcome variable was the annual rate of decline of absolute change in FVC (post- bronchodilator) from Baseline to Week 52. Baseline FVC was defined as the maximum acceptable FVC measurements obtained during the screening visit. The FVC at Week 52 was defined as the mean of the maximum acceptable FVC measurements obtained on two separate days at the Week 52 visit (Week 52A and Week 52B).

The analysis of the primary endpoint was a random coefficients linear regression model, with an absolute change in FVC as the outcome variable assuming linear decline in lung function over time. The model included random coefficients for intercept and slope and fixed effect terms for treatment, sex, age, and height. The statistical model is described in the statistical analysis plan (SAP) as follows:

$$Y_{itk} = (\alpha + a_i) + (\gamma + \beta_s T_k + g_i)\, t + \beta_g \text{Gender}_i + \beta_a \text{Age}_i + \beta_h \text{Height}_i + \varepsilon_{it}$$

- Y_{itk} is the value measured for ith patient at time t in treatment group k
- T_k =0 if patient in Placebo group and T_k=1 is patient in Nintedanib 150mg bid
- β_s is the effect of Nintedanib 150mg bid on the slope
- α and γ are elements of the intercept and slope respectively a_i and g_i are random specific components of the intercept and slope for the ith patient
- β_g , β_a and β_h are patient specific demographics' coefficients
- Gender_i, Age_i and Height_i are the gender (Male as the class of reference), baseline age [years] and baseline height [cm] for the ith patient
- ε_{it} is the random error for ith patient at time t
- a_i and g_i are assumed to be normally distributed with mean 0 and arbitrary covariance matrix
- ε_{it} are assumed to be independent and normally distributed with mean 0 and variance σ_{ε^2}
- Within patient errors follow a random coefficient regression model with random effect for intercept and slope
- An unstructured variance-covariance structure will be used to model the within patient measurements
- The variance-covariance matrix, modeled to estimate the inter-individual variability is considered to have a Variance-Components structure

In the analysis of the primary endpoint, missing data were not imputed and assume to be missing-at-random (MAR). In other words, the analysis was conducted on observed cases (OC). This assumption initially appeared to me as unacceptable since missing data from treatment discontinuation were not randomly occurring. However, the statistical model used by the applicant assumes a linear decline in lung function over time and implicitly imputes missing data based on individual's estimated rate of worsening of lung function prior to treatment discontinuation, similar to linear extrapolation. Our concern regarding

the MAR assumption was conveyed to the applicant in the EOP2 meeting and we recommended sensitivity analyses to assess impact of missing data. Subsequently, the applicant proposed the following sensitivity analyses regarding missing data in the SAP:

> c) To investigate the potential effect of missing data, patients will be classified into different patterns depending on the availability of data:
>
> > o Patients with a 52 week FVC value (or a value after week 52 but before follow-up visit):
> >
> > > 1. those who received trial drug until 52 weeks (defined as patients who did not prematurely discontinue the trial medication as according to the "end of trial medication" page of the CRF)
> > >
> > > 2. those who prematurely discontinued trial drug (as per information given on the "end of trial medication" page of the CRF) but who were followed up until week 52
> >
> > o Patients without a 52 week FVC value:
> >
> > > 3. those who were alive at 52 weeks (based on "vital status" page of the CRF, and no fatal AE recorded on the "adverse event" page of the CRF)
> > >
> > > 4. those who died before 52 weeks (based on "vital status" page of the CRF, or fatal AEs recorded on the "adverse event" page of the CRF)
>
> These four patterns described under c) will be used in sensitivity analyses to estimate the treatment effect under differing assumptions regarding the persistence of efficacy post withdrawal of randomised treatment, using multiple imputation.
>
> Non-monotone missing data or missing data at other visits before week 52 will not be imputed. Multiple imputation will be used to handle missing data at week 52. The imputation model will be similar to the statistical model of the primary analysis. For the imputation of week 52 data in pattern 3, the imputation model will be run on the subset of patients who prematurely discontinued trial drug but have been followed up for FVC measurements at week 52 (pattern 2) for sensitivity analyses 1 and 2. The slope (SE) estimates of both treatment groups will be used in sensitivity analysis 1. This approach is considered appropriate since the reasons for trial drug termination appear to be similar in pattern 2 and 3 (discontinuation mainly due to AEs), based on a blinded assessment of the missing data patterns. Sensitivity analysis 1 corresponds to the assumption that in pattern 3 patients, the treatment effect would have persisted in the same manner as for pattern 2 patients after trial drug discontinuation. In sensitivity analysis 2, only the placebo slope will be used to impute the missing week 52 data in all patients regardless of the randomised treatment group, corresponding to the assumption that in pattern 3 patients, the treatment effect in patients randomised to Nintedanib 150 mg does not persist after the discontinuation of trial drug but that instead all patients in pattern 3 would have had slopes like the placebo patients in pattern 2. For sensitivity analyses 1 and 2, pattern 2 patients are used as the basis for multiple imputations but since the number of patients in that pattern may be small, a third sensitivity analysis will be performed to confirm the robustness of the primary analysis results. Sensitivity analysis 3 makes similar assumptions as those in sensitivity analysis 2, except that the imputation model will be run on all patients randomised in the placebo group and hence using the placebo slope (SE) estimated in the primary analysis to impute missing week 52 data in all patients. This corresponds to the assumption that in pattern 3, all patients would have had slopes similar to the slope estimated in all the placebo patients in the trial.
>
> Pattern 4 consists of patients who don't have week 52 data because they died before week 52. Assuming that deaths observed in the trial will likely be related to worsening of IPF, it seems reasonable to assume that the unobserved FVC values should on average be lower than those in patients who did not die prior to week 52. Therefore, the imputation model will be run on patients randomised in the placebo group (either in the subset of placebo patients included in pattern 2 for sensitivity analyses 1 and 2 or in all patients randomised to placebo in sensitivity analysis 3), but using a truncated distribution to impute FVC values at

week 52 as follows: If β represents the true slope with $f(\beta) \sim N(\beta, \sigma^2)$ where β and σ^2 are the placebo slope and SE estimates from either patients in pattern 2 or all placebo patients, then sampling for patients who died prior to 52 weeks is restricted to the interval $(-\infty,\beta]$ of the truncated distribution $f(\beta)/2$. In this way, it is guaranteed that, on average, the imputed FVC slope for patients who died is steeper than the average slope in patients who survived to week 52.

The key secondary efficacy outcome variables for each study were as follows:

- Time to first acute IPF exacerbation, defined as the first to occurrence of one of the following events:
  - Acute IPF Exacerbation by investigator's report
  - Unexplained worsening or development of dyspnea within 30 days
  - New diffuse pulmonary infiltrates on chest X-ray, and/or new HRCT parenchymal abnormalities with no pneumothorax or pleural effusion (new ground-glass opacities) since last visit
  - Exclusion of infection as per routine clinical practice and microbiological studies
  - Exclusion of alternative causes as per routine clinical practice, including the following:
    - Left heart failure
    - Pulmonary embolism
    - Identifiable cause of acute lung injury

- Change from baseline in SGRQ total score at Week 52

The log-rank test and Kaplan-Meier estimator were used to compare treatment groups for the time to first acute IPF exacerbation. The estimate of hazard ratio between two groups and its 95% confidence interval were obtained by Cox's proportional hazards regression model. The model included fixed effect terms for treatment, sex, age, and height.

A mixed model repeated measures (MMRM) was used for analysis of SGRQ total score change from baseline at Week 52. The model included a random effect for subject and fixed effect terms for treatment, visit, treatment-by-visit interaction, baseline, baseline-by-visit interaction, sex, age, and height.

The applicant proposed a hierarchical testing procedure to adjust for multiple endpoints. To test the key secondary endpoints, the primary endpoint must be statistically significant at the 5% level. If so, then the first secondary endpoint, time to first acute exacerbation is tested. If statistically significant at 5% level, then the next key secondary endpoint, SGRQ total score, is tested. There were no adjustments for the exploratory endpoints.

The exploratory efficacy outcome variables for each study were as follows:

- Time to death due to respiratory cause
- Time to death
- Time to death or lung transplant
- Shortness of Breath Questionnaire (SOBQ)
- Cough and Sputum Assessment Questionnaire (CASA-Q)

Similar statistical analyses as in the primary and key secondary endpoints analyses were conducted for those exploratory endpoints.

Sample Size Calculation

The primary efficacy analysis was adequately powered for evaluating the primary efficacy outcome variable for nintedanib versus the placebo group in both studies. Based on the applicant's sample size calculation, 194 patients in placebo group and 291 patients in nintedanib group would provide 90% power to detect a treatment difference of 100 mL in the absolute change in FVC between Baseline and Week 52, assuming a standard deviation of 300 mL at a significance level of 0.05.

Changes in the SAP

There were two amendments to the original SAP (June 11, 2013), Amendment 1 (September 11, 2013), and Amendment 2 (November 12, 2013). The applicant claimed that these amendments were made prior to unblinding and analyses of the efficacy data. Most of the changes were updated section reference, corrected efficacy analyses procedure, clarified wording to make SAP more complete and clear:

> The scope of sensitivity analysis has been clarified. Some sensitivity analysis was removed and multiple imputation approach was added as a sensitivity analysis following blinded assessment of missing data patterns and addressing the recommendations made by the FDA at Type C meeting. The interpretation of sensitivity analysis has also been further clarified. Details how to implement the multiple imputation approach have been added to Section 9.2.

Patient Disposition, Demographic and Baseline Characteristics

A total of 1066 patients (640 nintedanib and 426 placebo) were randomized (Table 2) and the majority (77%) of patients completed the 52 weeks of active treatment. The most common reason for discontinuations was adverse event. Compared to placebo, nintedanib treated patients had a higher percentage of dropouts due to an adverse event.

The disposition of patients is summarized in two ways. First I present the disposition for those subjects that discontinued study treatment but completed the study. Second, I present the disposition for those subjects that discontinued study treatment and withdrew from the study. Results are shown in Table 2.

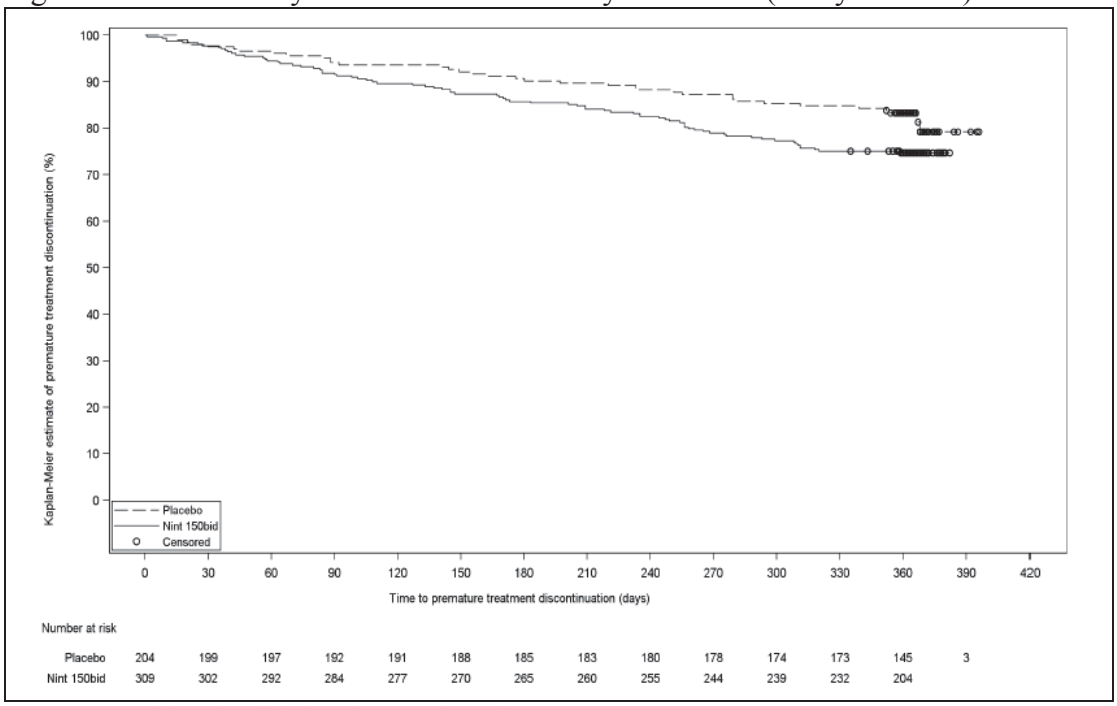
Table 2. Patients' Accountability, N (%) (All Randomized Patients)

	Study 1199.32 (N=515)		***Study 1199.34 (N=551)***	
	Nintedanib (n=309)	Placebo (n=206)	Nintedanib (n=331)	Placebo (n=220)
Received study treatment	309	204	329	219
Completed study treatment	231 (75)	168 (82)	251 (76)	175 (80)
Discontinued study treatment	78 (25)	36 (18)	78 (24)	44 (20)
Reason of early discontinuation of study treatment				
Adverse event	65 (21)	24 (12)	62 (19)	35 (16)
Non-compliant with protocol	2 (1)	3 (2)	2 (1)	1 (1)
Lost to follow-up	0	0	0	1 (1)
Patient refusal to continue taking trial medication	9 (3)	7 (3)	11 (3)	6 (3)
Other	2 (1)	2 (1)	3 (1)	1 (1)
Received treatment	309	204	329	219
Completed study	260 (84)	174 (85)	272 (83)	179 (82)
Discontinued study	49 (16)	30 (15)	57 (17)	40 (18)
Reason of withdrawal from the study				
Adverse event	25 (8)	15 (7)	42 (13)	30 (14)
Non-compliant with protocol	0	2 (1)	2 (1)	0
Lost to follow-up	0	0	2 (1)	1 (1)
Consent withdrawal, not due to adverse event	23 (7)	12 (6)	9 (3)	7 (3)
Other	1 (1)	1 (1)	2 (1)	2 (1)

Source: Excerpted from the Summary of Clinical Efficacy (page 51).

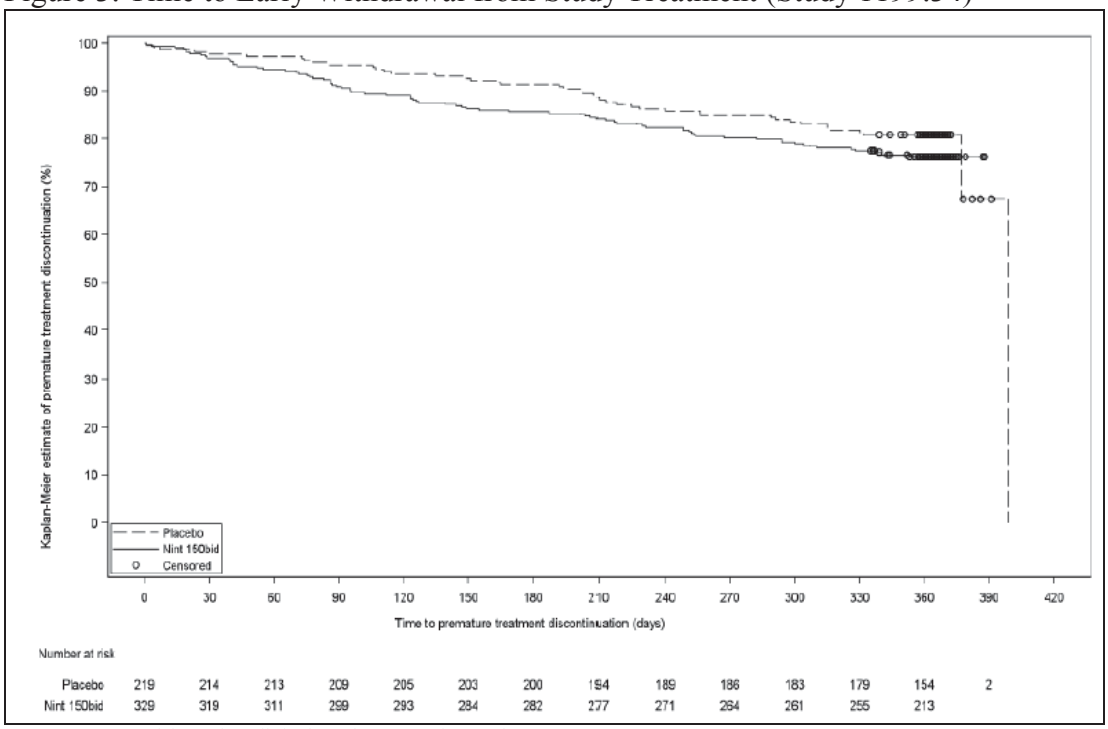
The survival curves for premature study drug discontinuations are presented in Figure 2 and Figure 3. The dropout rates were slightly higher in the nintedanib group compared to the placebo group.

Figure 2. Time to Early Withdrawal from Study Treatment (Study 1199.32)



Source: Excerpted from the Clinical Study Report for Study 1199.32 (page 97).

Figure 3. Time to Early Withdrawal from Study Treatment (Study 1199.34)



Source: Excerpted from the Clinical Study Report for Study 1199.34 (page 97).

In both studies, the demographic and baseline disease characteristics were generally well balanced and comparable between the treatment groups (Table 3). Overall, the mean age was 67 years. Majority of patients were Caucasian and approximately 80% of patients were male. The overall proportion of Asian patients was 30%; the proportion of Asian patients was higher in Study 1199.34 (39% vs. 21% in trial 1199.32). In Study 1199.32, mean baseline FVC in placebo group (2845 mL) was higher than that in nintedanib group (2757 mL). The imbalance was mainly due to lower mean baseline FVC among Asian nintedanib group (2586 mL) compared to those in other race-by-treatment groups (2803 mL – 2830 mL). However, in my sensitive analysis on FVC adjusting for the baseline FVC, I found that the apparent imbalance had no impact on the statistical significance.

Table 3. Patients' Demographic and Baseline Characteristics by Treatment, N (%)

	Study 1199.32 (N=515)		Study 1199.34 (N=551)	
Demographic parameter	Nintedanib (n=309)	Placebo (n=206)	Nintedanib (n=331)	Placebo (n=220)
Age at Randomization (yrs)				
Mean (SD)	67 (8.4)	67 (8.2)	66 (7.9)	67 (7.5)
Sex				
Male	251 (81)	163 (80)	256 (78)	171 (78)
Female	58 (19)	41 (20)	73 (22)	48 (22)
Race				
White	198 (64)	135 (66)	162 (49)	113 (52)
Black	0 (0)	0 (0)	2 (1)	0 (0)
Asian	66 (21)	41 (20)	128 (39)	87 (40)
Missing	45 (15)	28 (14)	37 (11)	19 (8)
Geographic region				
ROW	266 (86)	179 (87)	275 (83)	185 (84)
US	43 (14)	27 (13)	56 (17)	35 (16)
Time since IPF diagnosis (yrs)				
Mean (SD)	1.7 (1.4)	1.6 (1.4)	1.6 (1.3)	1.6 (1.3)
FVC (mL)				
Mean (SD)	2757 (735)	2845 (820)	2673 (776)	2619 (787)
SGRQ total score				
Mean (SD)	40 (18)	40 (18)	40 (20)	40 (19)

Source: Reviewer
Note: Patients randomized in French sites are shown under the 'Missing' category for race, because collection of race information is not allowed in France.

The average percentage of compliance to the study treatment was above 90% in both studies (Table 4). The median duration of treatment was close to 12 months in both studies while the mean duration was slightly above 10 months except for nintedanib group in Study 1199.34 which was 8.8 months.

Table 4. Study Treatment Compliance and Duration by Treatment

	Study 1199.32 (N=515)		Study 1199.34 (N=551)	
Treatment compliance	Nintedanib (n=309)	Placebo (n=206)	Nintedanib (n=331)	Placebo (n=220)
Patients who received any amount of of study treatment				
N (%)	309 (100)	204 (100)	329 (100)	219 (100)
Percent compliance per patient				
Mean (SD)	96 (7)	97 (7)	97 (6)	96 (6)
Median (Range)	98 (56-117)	99 (35-118)	98 (60-120)	99 (58-108)
N (%)				
<50%	0 (0)	1 (1)	0 (0)	0 (0)
50% to 80%	11 (4)	3 (2)	4 (1)	4 (2)
80% to 120%	284 (92)	198 (97)	316 (96)	208 (95)
Missing	14 (4)	2 (1)	9 (3)	7 (3)
Treatment duration in months				
Mean (SD)	10.3 (3.3)	10.9 (2.8)	8.8 (4.1)	10.6 (2.9)
Median (Range)	11.9 (0.0-12.5)	11.9 (0.5-13.0)	11.6 (0.0-12.7)	11.9 (0.0-13.1)

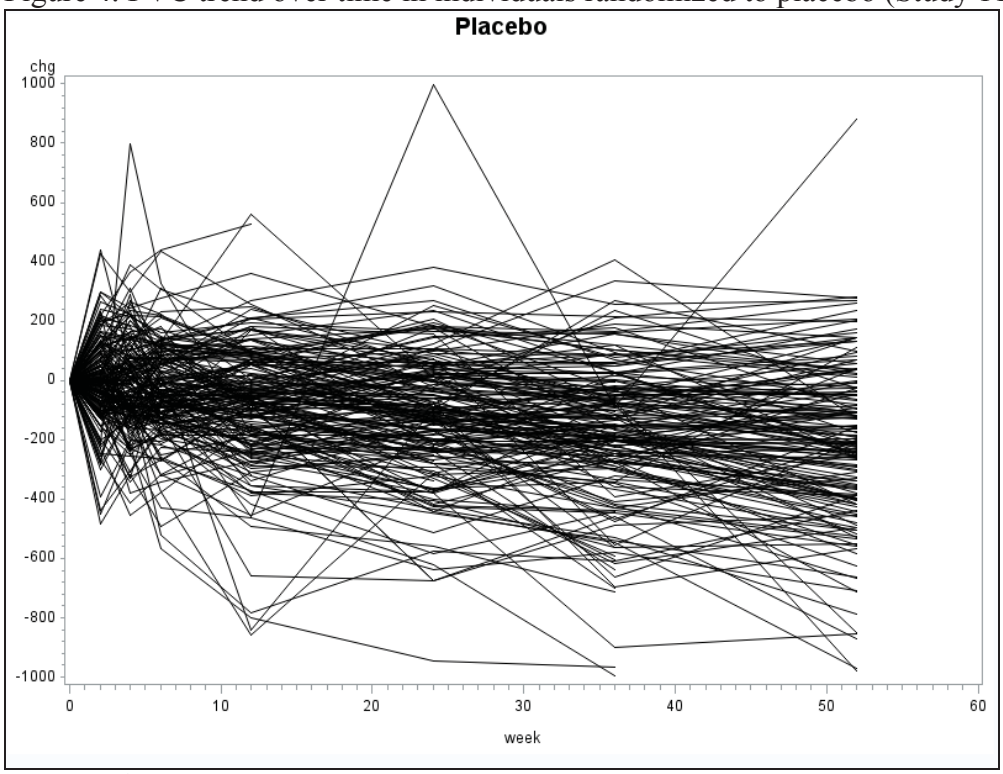
Source: Reviewer

Results and Conclusions

Primary Efficacy Endpoint – Annual rate of decline in FVC from Baseline to Week-52

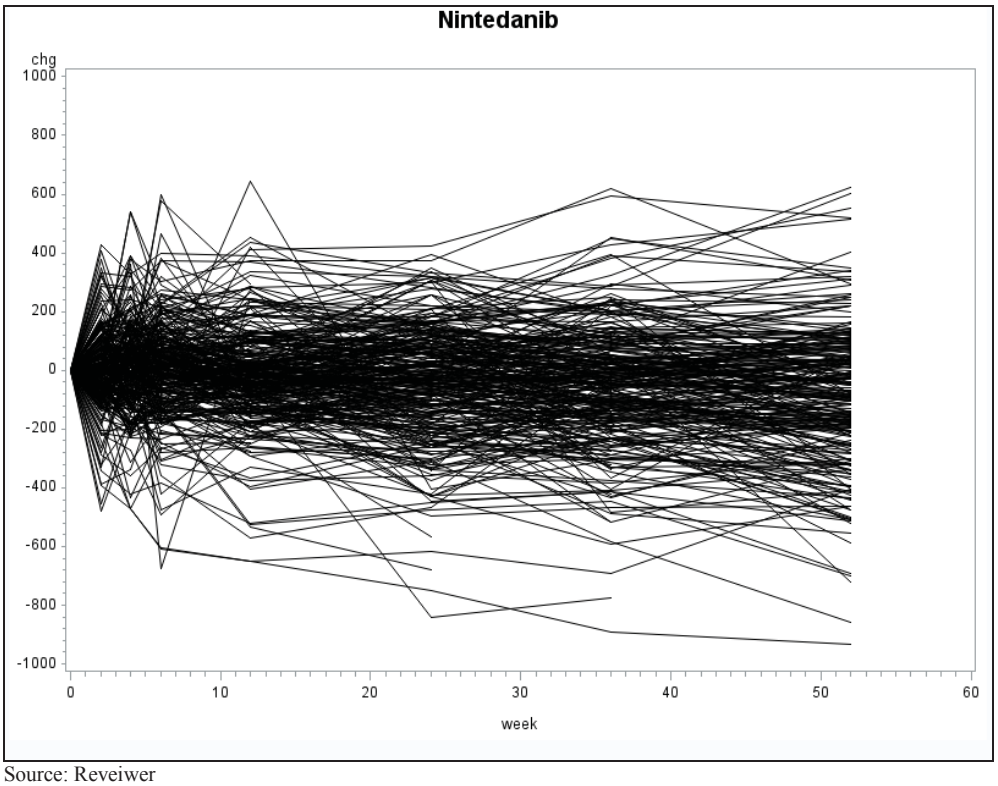
The following figures (4-6) describe the FVC change from baseline over time in each individual patient by treatment group. In Study 1199.32, majority of patients seem to experience decline in FVC although degree of decline appears slightly smaller in nintedanib group. In group mean graphs, the slope of decline in FVC of nintedanib group is smaller than the slope of placebo group (Figure 4). Similar interpretation can be drawn from the graphs generated from Study 1199.34 (results not shown).

Figure 4. FVC trend over time in individuals randomized to placebo (Study 1199.32)



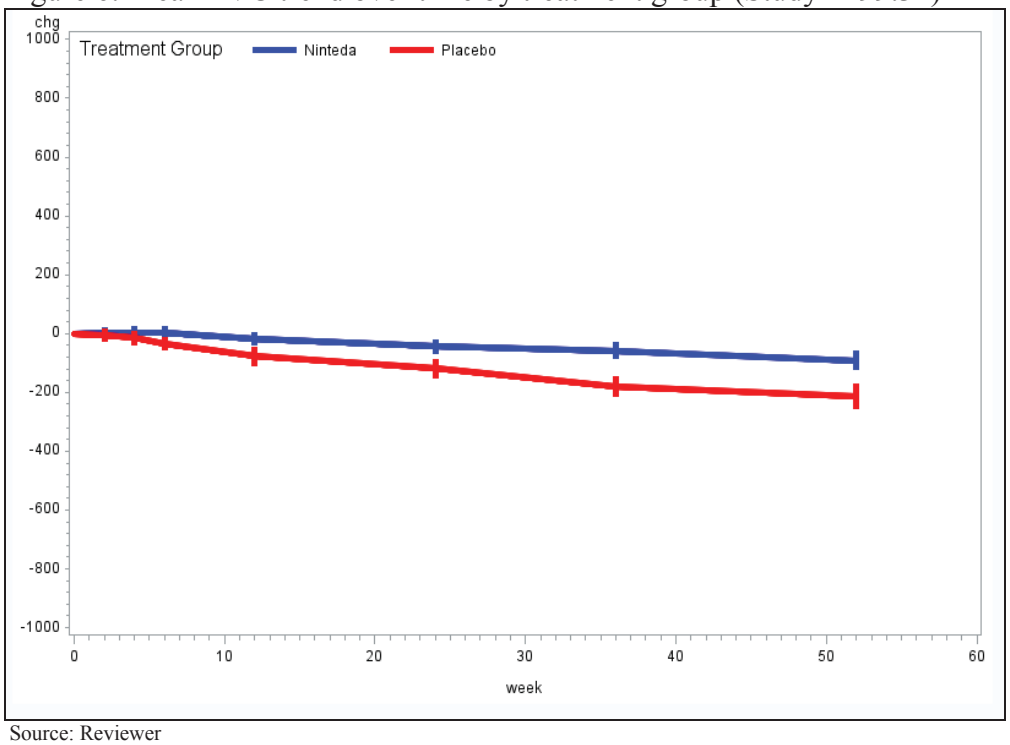
Source: Reveiwer

Figure 5. FVC trend over time in individuals randomized to nintedanib (Study 1199.32)



Source: Reveiwer

Figure 6. Mean FVC trend over time by treatment group (Study 1199.32)



Source: Reviewer

The analysis of the primary endpoint was a random coefficient regression model without imputing for missing data. Although there was a non-ignorable amount of missing data due to treatment

dropouts, the estimate of the slope for decline from the analysis was reasonably conservative since the analysis model assumed a linear decline of FVC. Sensitivity analyses as specified in the statistical analysis plan as Sensitivity analysis #1, #2, and #3 were conducted to examine the potential effect of missing data on the reliability of the primary analysis. I conducted a sensitivity analysis with imputation for missing data with mean of placebo completers using ANCOVA model with terms for sex, age, height and baseline score as covariate.

In Study 1199.32, no differences between the treatment groups in terms of proportion of patients with missing data were observed. A total of 436 patients (85.0% [84.8% nintedanib; 85.3% placebo]) had an FVC value at Week 52. Of these patients, 231/262 patients in the nintedanib group and 163/174 patients in the placebo group received trial medication until Week 52. The remaining 31/262 patients in the nintedanib group and 11/174 patients in the placebo prematurely discontinued treatment but completed the planned observation time until Week 52.
At Week 52, 77 patients (15.0% [15.2% nintedanib; 14.7% placebo]) had a missing FVC value. Of these patients, 14/47 patients in the nintedanib group and 13/30 patients in the placebo group had missing data because they died before Week 52. Of all the patients with missing data at Week 52, 27/47 patients in the nintedanib arm and 20/30 patients in the placebo arm had FVC data up to Week 24 or Week 36. Sensitivity analyses carried out by the applicant to assess the robustness included analyses using 3 different scenarios of multiple imputations for missing Week 52 data.

Patients receiving nintedanib had a smaller mean decline from Baseline in FVC compared to those receiving placebo at Week 52 ($p < 0.001$, random coefficients regression model) in Study 1199.32 (Table 5). An estimated absolute difference was 125 mL (i.e. -115 – -240 = 125 mL FVC) between the two treatment groups. Sensitivity analyses with respect to missing data by the applicant and me gave consistent results from the primary analyses.

Table 5. Analyses on Annual Rate of Decline in FVC from Baseline to Week 52 (Study 1199.32)

Treatment	N	LS Mean	Difference vs. PBO		
			LS Mean Difference	95% CI	P-value
Applicant's Primary Analysis: Treated Set (Random Coefficient Regression)					
Nintedanib	309	-115	125	(78, 173)	<0.001
Placebo	204	-240			
Applicant's Sensitivity Analysis #1: Treated Set (Random Coefficient Regression with MI)					
Nintedanib	309	-121	120	(76, 165)	<0.001
Placebo	204	-241			
Applicant's Sensitivity Analysis #2: Treated Set (Random Coefficient Regression with MI)					
Nintedanib	309	-127	115	(70, 160)	<0.001
Placebo	204	-242			
Applicant's Sensitivity Analysis #3: Treated Set (Random Coefficient Regression with MI)					
Nintedanib	309	-128	114	(69, 159)	<0.001
Placebo	204	-242			
My Sensitivity Analysis: Treated Set (ANCOVA with Placebo Mean Imputation)					
Nintedanib	309	-101	88	(45, 131)	<0.001
Placebo	204	-189			
My Sensitivity Analysis: Treated Set (Rank ANCOVA with Lowest Rank Imputation)					
Nintedanib	309	--	--	--	<0.001
Placebo	204	---			

Source: Reviewer

A similar significant result was found in Study 1199.34 (Table 6). An estimated absolute difference was 94 mL (i.e. -113 – -207 = 125 mL FVC) between the two treatment groups. Again, sensitivity analyses with respect to missing data by the applicant and me gave consistent results from the primary analyses.

Table 6. Analyses on Annual Rate of Decline in FVC from Baseline to Week 52 (Study 1199.34)

Treatment	N	LS Mean	Difference vs. PBO		
			LS Mean Difference	95% CI	P-value
Applicant's Primary Analysis: Treated Set (Random Coefficient Regression)					
Nintedanib	329	-114	94	(45, 143)	<0.001
Placebo	219	-207			
Applicant's Sensitivity Analysis #1: Treated Set (Random Coefficient Regression with MI)					
Nintedanib	329	-136	101	(52, 150)	<0.001
Placebo	219	-237			
Applicant's Sensitivity Analysis #2: Treated Set (Random Coefficient Regression with MI)					
Nintedanib	329	-154	83	(33, 133)	0.001
Placebo	219	-237			
Applicant's Sensitivity Analysis #3: Treated Set (Random Coefficient Regression with MI)					
Nintedanib	329	-124	83	(38, 129)	<0.001
Placebo	219	-208			
My Sensitivity Analysis: Treated Set (ANCOVA with Placebo Mean Imputation)					
Nintedanib	329	-112	95	(50, 139)	<0.001
Placebo	219	-206			
My Sensitivity Analysis: Treated Set (Rank ANCOVA with Lowest Rank Imputation)					
Nintedanib	329	--	--	--	0.005
Placebo	219	--			

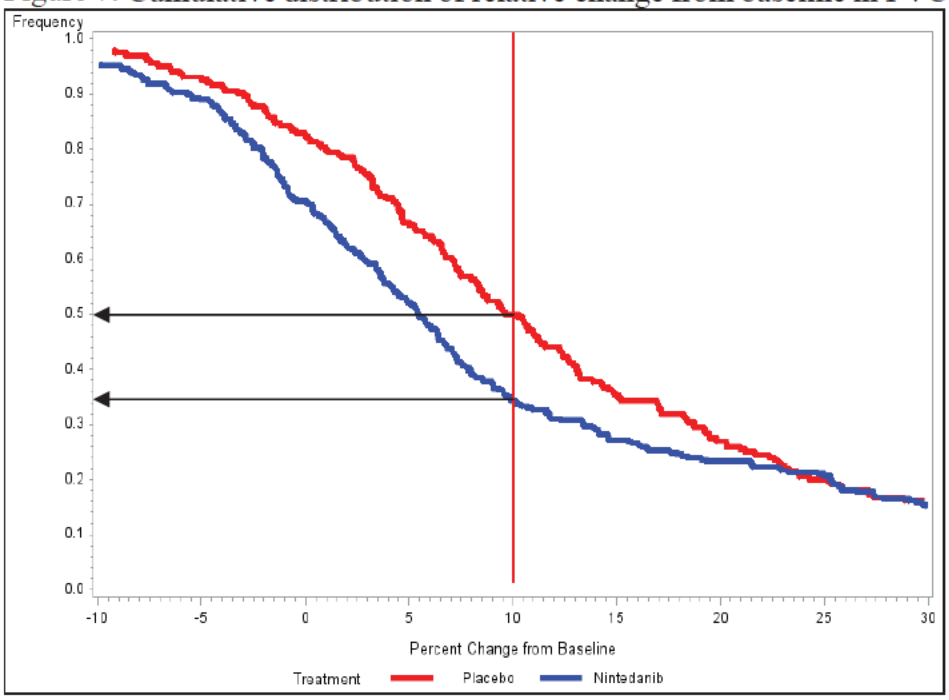
Source: Reviewer

I also conducted a continuous responder analysis. In each study, continuous responder curves for each treatment arm were plotted. In these plots, all patients who drop out from treatment due to any reason are considered non-responders (i.e. highest decline in FVC). Note that these figures were created to provide a visual display of the relative benefit of nintedanib across the entire range of response at Week 52. The x-axis shows the relative decline in FVC from baseline (or worsening) at Week 52 and the y-axis show the corresponding percentage of patients achieving that level of FVC decline or greater. The positive treatment effect of nintedanib was demonstrated by a consistent separation of the curves across all levels of response in Study 1199.32. As an example, only 35% of nintedanib-treated patients have 10% or greater decline in FVC compared to 50% of placebo-treated patients (Figure 7). If we interpret '10% or less decline' as response, then proportion of responders was 65% and 50% in nintedanib group and placebo group, respectively. With the same definition, proportion of responders was 64% and 54% in nintedanib group and placebo group, respectively in Study 1199.34 (Figure 8).

In consultation with the clinical team, a cut-off point of at least a 10% relative decline in FVC was chosen to perform a responder analysis. The results from this responder analysis confirmed the

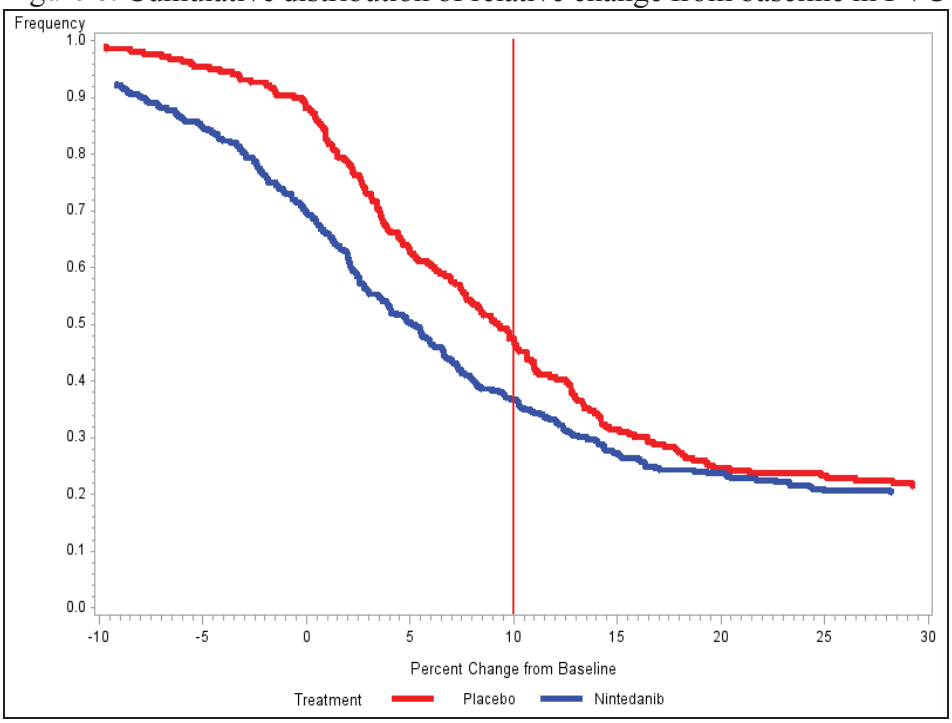
results of the primary analysis. Nintedanib demonstrated a benefit in reducing the decline of lung function in studies 1199.32 and 1199.34. The applicant also conducted responder analyses using %Predicted FVC considering two thresholds, at least a 5% decline and at least a 10% decline. Dropouts were considered non-responders. With the 10% threshold in Study 1199.32, the proportion of responders was 71% and 57% in the nintedanib group and placebo group, respectively. For Study 1199.34, the proportion of responders was 70% and 64% in the nintedanib group and placebo group, respectively. Using a logistic regression analysis, the differences noted above were only significant for Study 1199.32. However, when the 5% threshold was considered, statistical significance in favor of nintedanib was observed in both studies. Results are shown in Tables 17 and 18 in the appendix.

Figure 7. Cumulative distribution of relative change from baseline in FVC (Study 1199.32)



Source: Reviewer

Figure 8. Cumulative distribution of relative change from baseline in FVC (Study 1199.34)



Source: Reviewer

I conducted additional analyses on the FVC (mL) with a rank ANCOVA model with terms for sex, age, height and ranked baseline score as covariate assigning the worst rank for dropouts from treatment to assess the impact of distributional assumption on the results from primary analyses.

The analyses results were consistent with the significant results from the primary analyses in both studies ($p<0.001$ for Study 1199.32 and $p=0.005$ for Study 1199.34).

In summary, the two phase 3 studies in patients with IPF, showed statistically significant evidence in favor of nintedanib on the change in lung function (primary efficacy endpoint). In both studies, several secondary analyses were conducted on the primary efficacy endpoint to assess the robustness of the primary analysis. Although the magnitude of treatment effects varies depending on the methods of imputation and the statistical approaches used, the conclusions from these analyses were consistent.

Secondary Efficacy Endpoints

I was able to confirm the results of the applicant's analyses of the secondary endpoints. A review of the two pre-specified secondary efficacy endpoints in the hierarchical order for multiple testing is described in the next subsections for each individual study. Also a review on all-cause mortality endpoint is presented for individual study data as well as pooled data.

Key Secondary Endpoints - The Time to First Acute IPF Exacerbation

The applicant's results of the time to first acute IPF exacerbation analysis are summarized in Table 7 and Figures 9 and 10. Kaplan-Meier estimates were used to summarize time to first acute exacerbation, and treatment differences were analyzed using the log-rank test. The hazard ratio (HR) was determined based on the Cox proportional hazard model, adjusted for sex, baseline height, and baseline age, to estimate the magnitude of the effect. Censoring was applied at 372 days after randomization. Patients for whom no exacerbation event was reported within 373 days (included) of randomization were censored at 373 days or last contact date, whichever occurred first.

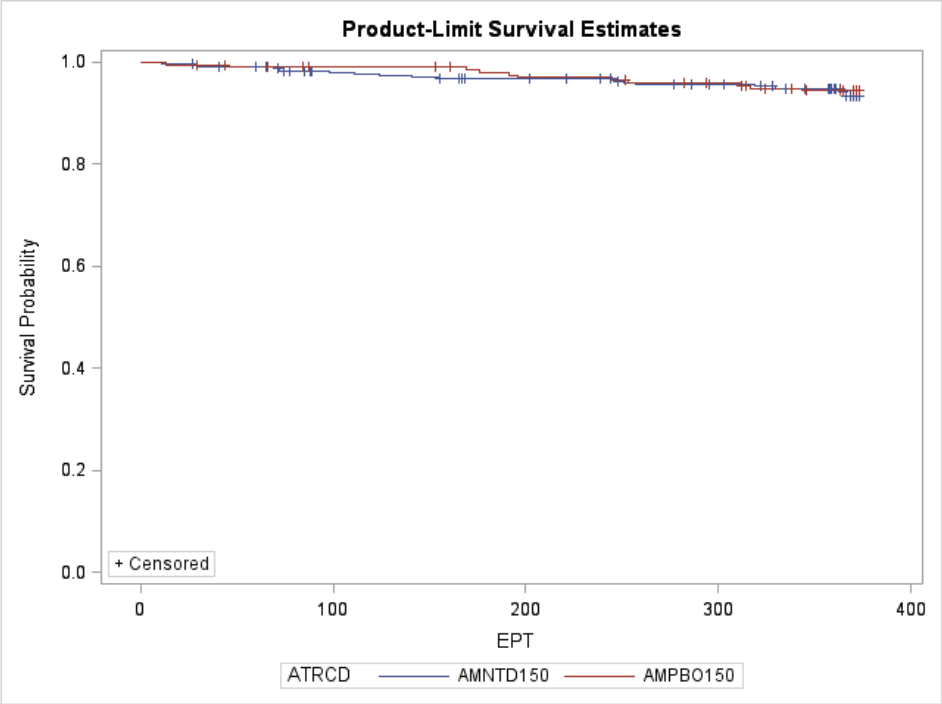
In Study 1199.34, treatment with nintedanib resulted in a lower proportion of patients with at least one acute exacerbation than treatment with placebo, 3.6% vs. 9.6%, respectively. Treatment with nintedanib was associated with a 62% relative reduction of the risk of acute exacerbation compared to placebo (HR [95% CI]: 0.38 [0.19–0.77], p=0.005). There was also evidence of a treatment effect of nintedanib that began at approximately Week 12 and extended toward Week 52 (Figure 10). This evidence was not seen in Study 1199.32 (Figure 9). Since replicated evidence for this endpoint was not demonstrated in the two phase 3 trials, I assessed the significance of this endpoint in the phase 2 study, 1199.30. These results are presented in section 3.2.2.

Table 7. Survival Analysis on Time to First Acute Exacerbation over 52 weeks

	Nintedanib	***Placebo***	***Hazard Ratio (95% CI)*** [b]
	N of Event (%)	***N of Event (%)***	***p-value*** [a]
Study 1199.32			
N of Randomized	309	204	
Acute exacerbation	19 (6.1)	11 (5.4)	1.15 (0.54, 2.42), 0.673
Study 1199.34			
N of Randomized	329	219	
Acute exacerbation	12 (3.6)	21 (9.6)	0.38 (0.19, 0.77), 0.005

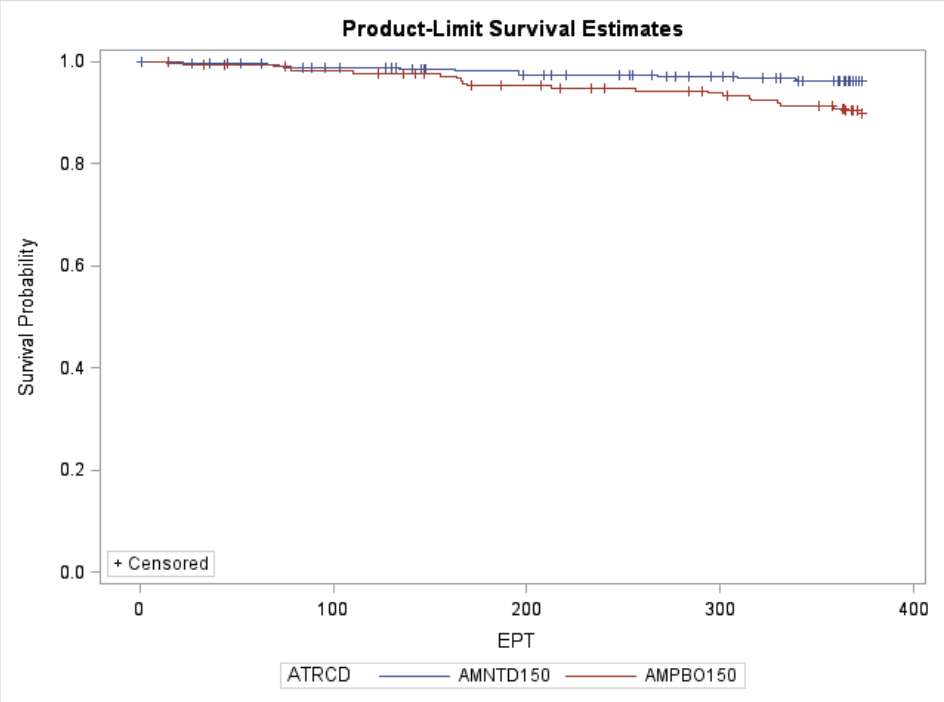
Source: Reviewer
[a] p-value was based on the log-rank test.
[b] Hazard ratio was based on the Cox proportional hazard model with terms for treatment, sex, age, and height.

Figure 9. Kaplan-Meier Curve of Time to First Acute Exacerbation over 52 weeks (Study 1199.32)



Source: Reviewer

Figure 10. Kaplan-Meier Curve of Time to First Acute Exacerbation over 52 weeks (Study 1199.34)



Source: Reviewer

Key Secondary Endpoints - The Change from baseline in SGRQ total score

The results from the analyses of the mean change from baseline in SGRQ total score are summarized in Table 8. The endpoint was analyzed using MMRM model with fixed effects for treatment, visit, treatment-by-visit, baseline SGRQ total score, baseline SGRQ total score-by-visit and random effect for patient.

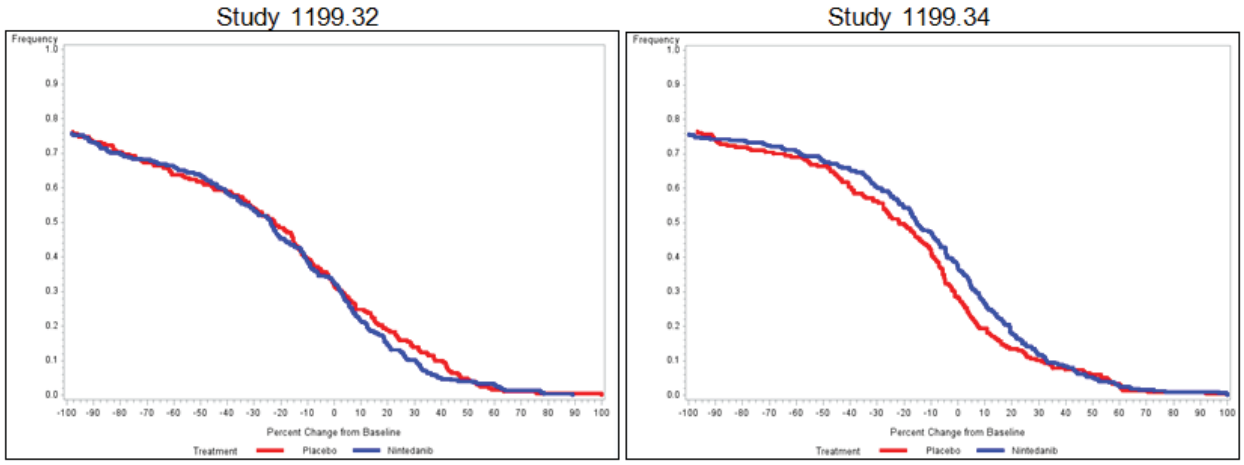
In Study 1199.34, the mean change in SGRQ total score in patients treated with nintedanib was significantly lower compared to patients treated with placebo (2.8 vs. 5.5, respectively; difference of -2.7, p=0.020). This effect was not observed in Study 1199.32. My sensitivity analyses with ANCOVA model with terms for treatment and baseline SGRQ total score after imputing missing data with mean of placebo completers were consistent with results from the applicant's pre-specified MMRM analyses. Also my cumulative distribution curves with worst score imputation for missing data showed separation of curves in Study 1199.34, but not in Study 1199.32 (Figure 11). Again since replicated evidence of efficacy for this endpoint was not demonstrated in the two phase 3 studies, I considered the results from the, phase 2 study, 1199.30. These results are presented in section 3.2.2.

Table 8. Change from Baseline in SGRQ total score at Week 52

	Study 1199.32 (N=515)		***Study 1199.34 (N=551)***	
	Nintedanib (n=309)	Placebo (n=204)	Nintedanib (n=329)	Placebo (n=219)
Applicant's MMRM analysis				
N	289	202	320	213
LSMEAN (SE)	4.3 (0.8)	4.4 (1.0)	2.8 (0.7)	5.5 (0.9)
vs. Placebo	-0.1 (1.3)		-2.7 (1.2)	
95% CI	(-2.5, 2.4)		(-5.0, -0.4)	
p-value	0.966		0.020	
My ANCOVA analysis with placebo mean imputation				
N	309	204	329	219
LSMEAN (SE)	3.3 (1.0)	3.2 (1.1)	1.5 (0.9)	4.1 (1.1)
vs. Placebo	-0.1 (1.3)		-2.6 (1.2)	
95% CI	(-2.5, 2.4)		(-5.0, -0.2)	
p-value	0.981		0.034	

Source: Reviewer

Figure 11. Cumulative distribution of relative change from baseline in SGRQ total score



Source: Reviewer

Absolute Change from Baseline to Week 52 in FVC

In Study 1199.32, the adjusted mean absolute change from baseline to Week 52 in FVC was lower in the nintedanib group (-95 mL) than in the placebo group (-205 mL). The adjusted mean difference between the treatment groups was 110 mL (95% CI: 71, 149) and was statistically significant with p<0.001. In Study 1199.34, the adjusted mean absolute change from baseline to Week 52 in FVC was lower in the nintedanib group (-95 mL) than in the placebo group (-205 mL). The adjusted mean difference between the treatment groups was 110 mL (95% CI: 71, 149) and was statistically significant with p<0.001.

Absolute Change from Baseline to Week 52 in %Predicted FVC

In Study 1199.32, the adjusted mean absolute change from baseline to Week 52 in FVC% predicted was lower in the nintedanib group (-2.8%) than in the placebo group (-6.0%). The adjusted mean difference between the treatment groups was 3.2% (95% CI: 2.1, 4.3) and was statistically significant with p<0.001. In Study 1199.34, the adjusted mean absolute change from baseline to Week 52 in FVC% predicted was lower in the nintedanib group (-3.1%) than in the placebo group (-6.2%). The adjusted mean difference between the treatment groups was 3.1% (95% CI: 1.9, 4.3) and was statistically significant with p<0.001.

All-cause mortality

As IPF is a chronic progressive disease with survival estimated to be from 3 to 5 years following diagnosis, mortality is the ideal primary efficacy variable in IPF clinical trials. During the December 2007, Pre-IND meeting, the main clinical concern raised by the Division was the primary efficacy variable, FVC. The Division noted that mortality is the ideal primary endpoint and FVC is not an established surrogate for mortality. Therefore it is unclear what would constitute a clinically meaningful outcome based on FVC. The Division further noted that efficacy would be assessed by the totality of the data, including secondary endpoints.

Because of the importance of this endpoint, the applicant's results on mortality are included

in this review. I also conducted an additional analyses using the results from the phase 3 studies and the phase 2 study.

The applicant conducted an analysis comparing all-cause mortality between the treatment groups in both studies. Kaplan-Meier estimates were used to summarize survival time up to the end of the study treatment period. Survival time is measured by time from randomization to death. Treatment differences were analyzed using the log-rank test. The hazard ratio (HR) was determined based on the Cox proportional hazard model, with sex, age, and height as factors. The results are displayed in Table 9. Although there was numerically smaller proportion of deaths in nintedanib group compared to placebo group, the difference was not statistically significant (Figure 12).

Table 9. Survival Analysis on All-Cause Mortality during the Treatment Period (All Treated Patients)

	Nintedanib	***Placebo***	***Hazard Ratio (95% CI)*** [c]
	N of Event (%)	***N of Event (%)***	***p-value*** [b]
Study 1199.32			
N of TS	309 [a]	204 [a]	
Death	13 (4.2)	13 (6.4)	0.63 (0.29, 1.36), 0.288
Censored	296 (95.8)	191 (93.6)	
Study 1199.34			
N of TS	329 [a]	219 [a]	0.74 (0.40, 1.35), 0.299
Death	22 (6.7)	20 (9.1)	
Censored	307 (93.3)	199 (90.9)	

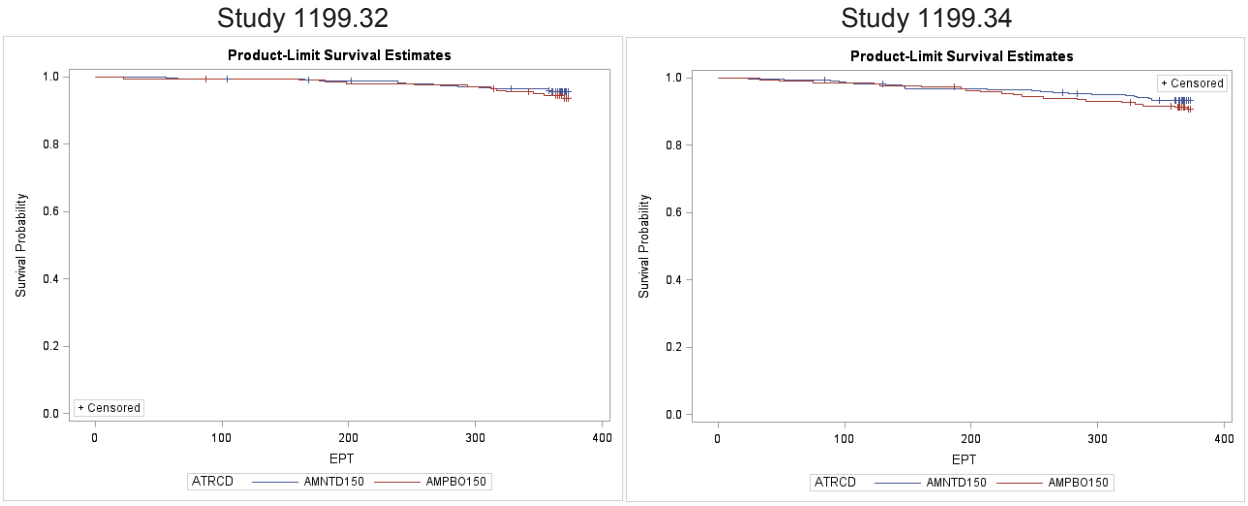
Source: Reviewer

[a] Based on occurrence of event, or censoring in the absence of the event. Time to event was the event date minus randomization date plus one. The censoring date was the last available contact date or time of rescue (if one occurred), or the end of the Treatment Period.

[b] p-value was based on the log-rank test comparing nintedanib with placebo.

[c] Hazard ratio was based on the Cox proportional hazard model with terms for treatment, sex, age, and height.

Figure 12. Kaplan-Meier Curve of Time to All-Cause Mortality during the Treatment Period



Source: Reviewer

Since both studies were not powered for the survival endpoint, a post-hoc analysis pooling the two studies was conducted to increase the power to detect difference if any. However, the results on pooled data still did not reach a statistical significance (6%, 35/638 vs. 8%, 33/423; HR [95% CI]: 0.70 [0.43, 1.12], p=0.140). Another analysis integrating the three studies, 1199.32, 1199.34, and 1199.30 was conducted and is presented in Section 3.2.2.

3.2.2 Study 1199.30

The objective of this study was to evaluate the efficacy and safety of 4 dose strategies of nintedanib treatment for 52 weeks compared with placebo in patients with IPF. The 4 doses of nintedanib were 50 mg qd, 50 mg bid, 100 mg bid, and 150 mg bid. For this study, I focus on two treatment groups, nintedanib 150 mg bid and placebo.

Study Design and Endpoints

This study consisted of a screening period, a study treatment period, and a final follow-up visit. After the screening visit, if the patient complies with all inclusion and exclusion criteria, randomization will be performed by phone or Internet. Patients will then enter the treatment phase for 52 weeks. Nine visits (visit 2 to 9 + follow up) are planned within one year of treatment, and intermediate lab tests are planned when the interval between two visits increases.

The diagnosis of IPF were confirmed by central review of high resolution computerized tomography (HRCT) and, if available, surgical lung biopsy. Patients were required to be aged≥40, diagnosed with IPF within last 5 years, and to have %Predicted FVC ≥ 50% and %Predicted carbon monoxide diffusing capacity (DLCO) 30% to 79%.

Spirometry measurements, including FVC and FEV_1 were to be assessed at Screening, Day 1

(before randomization), at Week 2, 4, 6, 12, 24, 36, 52 (or End of Treatment), and the Final Follow-up visit. At each visit, three FVC values were collected before and after bronchodilator, respectively, until maximum acceptable FVC value was chosen.

The primary efficacy outcome variable was the annual rate of decline of absolute change in FVC (post- bronchodilator) from Baseline to Week 52. Baseline FVC was defined as the maximum acceptable FVC measurements obtained during the screening visit. The FVC at Week 52 was defined as the mean of the maximum acceptable FVC measurements obtained on two separate days at the Week 52 visit (Week 52A and Week 52B).

Statistical methodologies
The analysis of the primary endpoint was the same random coefficients regression model as used for the two phase 3 studies, 1199.32 and 1199.34. The outcome variable was the absolute change in FVC and assumed a linear decline in lung function over time. The model included random coefficients for intercept and slope and fixed effect terms for sex, age, and height. The primary analysis population was the Intent-to-Teat (ITT) patient population (all randomized patients regardless of actually received treatment). Missing data were not imputed and assumed a MAR assumption. In other words, the analysis was conducted on observed cases (OC). As previously stated, this assumption initially appeared to me unacceptable since missing data from treatment discontinuation were not randomly occurring. However, the statistical model assumes a linear decline in lung function over time and implicitly imputes missing data similar to linear extrapolation based on individual's estimated rate of worsening of lung function prior to treatment discontinuation. The applicant proposed a sensitivity analysis with retrieved dropouts.

The secondary efficacy outcome variables were as follows:

- Change from baseline in %Predicted FVC at Week 52
- SGRQ
- Time to First Exacerbation of IPF
- Survival (all causes of death, and lung-transplant free) at 6 and 12 months
- SpO2 (oxygen saturation) at rest
- DLCO
- 6-Minute Walk Test

The applicant proposed a closed testing procedure to adjust for multiple comparisons between each dose group and placebo group for the primary endpoint only. They did not adjust for secondary endpoints.

<u>Sample Size Calculation</u>
Based on the applicant's sample size calculation, 400 patients randomized to 4 nintedanib dose groups and placebo in 1:1:1:1:1 ratio were expected to provide 80% power to detect a treatment difference of 100 mL in the absolute change in FVC between Baseline and Week 52, assuming a standard deviation of 200 mL at a significance level of 0.05.

Patient Disposition, Demographic and Baseline Characteristics

The focus of the review on this study will be on two treatment groups, nintedanib 150 mg bid and placebo.

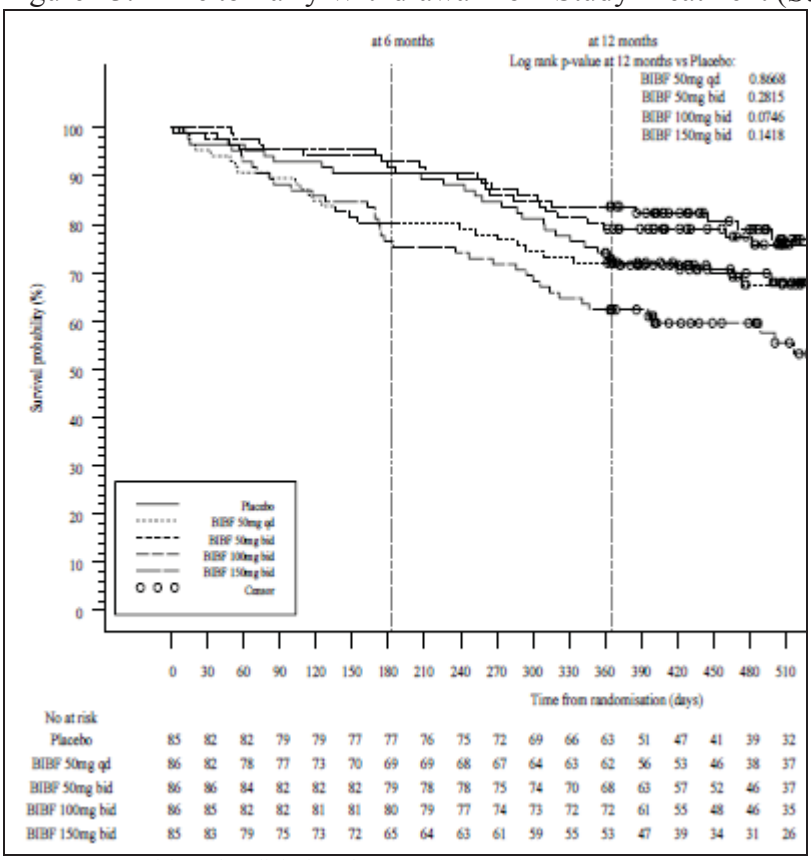
A total of 432 patients (345 nintedanib and 87 placebo) were randomized (Table 10) and the majority (74%) of patients completed the 52 weeks of active treatment. The most common reason for discontinuations was adverse event. Compared to placebo, nintedanib 150mg bid-treated patients had a higher percentage of dropouts due to adverse event and also discontinued study drug early as shown in the survival curves for premature study drug discontinuations (Figure 13).

Table 10. Patients' Accountability, N (%) (All Randomized Patients (Study 1199.30))

	Nintedanib 50mg qd (n=87)	Nintedanib 50mg bid (n=86)	Nintedanib 100mg bid (n=86)	Nintedanib 150mg bid (n=86)	Placebo (n=87)
Received study treatment	86	86	86	85	85
Completed study treatment	62 (72)	68 (79)	72 (84)	53 (62)	61 (72)
Discontinued study treatment	24 (28)	18 (21)	14 (16)	32 (38)	24 (28)
Reason of early discontinuation of study treatment					
Adverse event	20 (23)	15 (17)	13 (15)	27 (32)	21 (25)
Non-compliant with protocol	1 (1)	0	1 (1)	0	1 (1)
Lost to follow-up	0	0	0	0	0
Other	1 (1)	2 (2)	0	1 (1)	0

Source: Excerpted from the Clinical Study Report (page 116).

Figure 13. Time to Early Withdrawal from Study Treatment (Study 1199.30)



Source: Excerpted from the Clinical Study Report (page 119).

The demographic and baseline disease characteristics were generally well balanced and comparable between the treatment groups (Table 11). Overall, the mean age was 65 years. Majority of patients were Caucasian and approximately 75% of patients were male. The overall proportion of Asian patients was 20%. On average, patients had 1.2 year history of IPF after diagnosis.

Table 11. Patients' Demographic and Baseline Characteristics by Treatment, N (%) (Study 1199.30)

Demographic parameter	Nintedanib 50mg qd (n=86)	Nintedanib 50mg bid (n=86)	Nintedanib 100mg bid (n=86)	Nintedanib 150mg bid (n=85)	Placebo (n=85)
Age at Randomization (yrs)					
Mean (SD)	65 (9.4)	65 (8.5)	66 (7.9)	67 (7.5)	65 (8.6)
Sex					
Male	65 (76)	62 (72)	65 (76)	65 (77)	63 (74)
Female	21 (24)	24 (28)	21 (24)	20 (23)	22 (26)
Race					
White	68 (79)	72 (84)	72 (84)	61 (72)	65 (77)
Asian	18 (21)	14 (16)	14 (16)	24 (28)	20 (24)
Time since IPF diagnosis (yrs)					
Mean (SD)	1.4 (1.3)	1.1 (1.2)	1.2 (1.2)	1.0 (1.2)	1.4 (1.5)
FVC (L)					
Mean (SD)	2.8 (0.8)	2.7 (0.7)	2.9 (0.8)	2.7 (0.8)	2.8 (0.8)
SGRQ total score					
Mean (SD)	44 (18)	43 (17)	44 (17)	40 (18)	41 918)

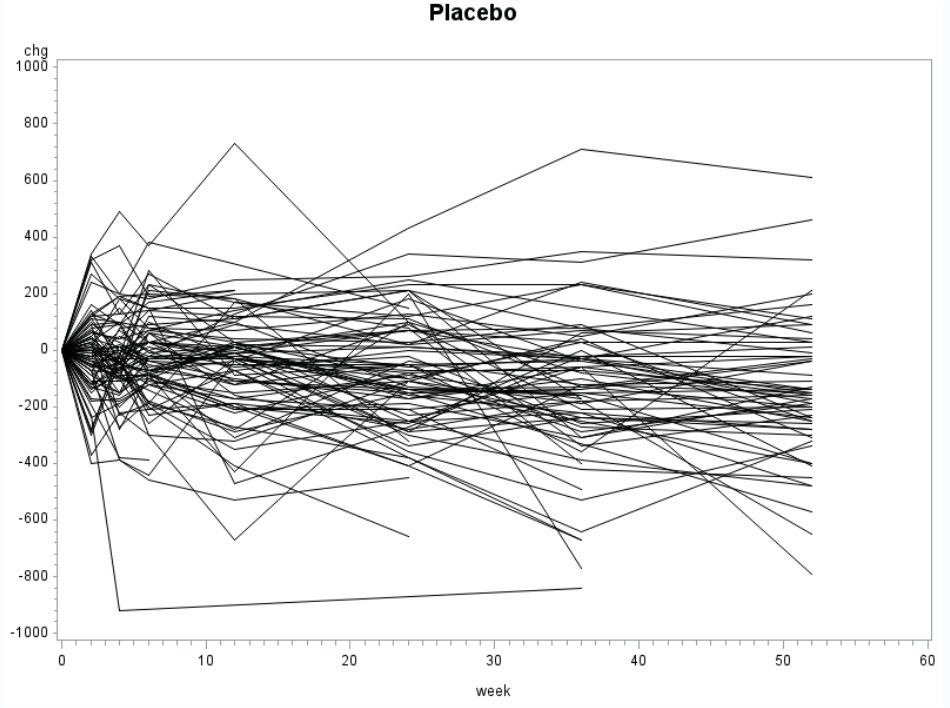
Source: Reviewer

Results and Conclusions

Primary Efficacy Endpoint – Annual rate of decline in FVC from Baseline to Week-52

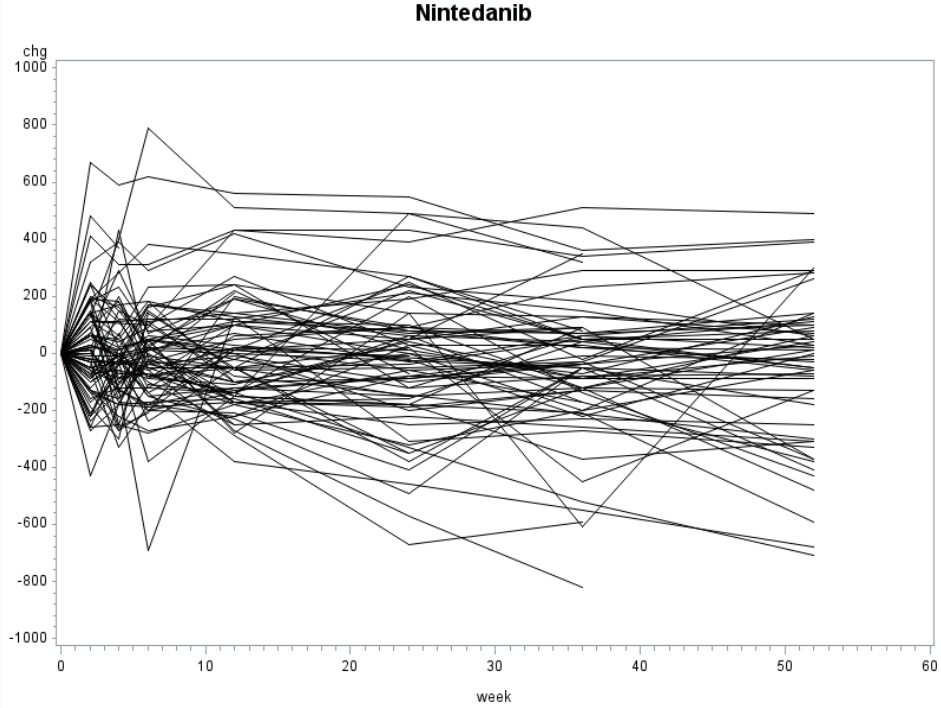
In this section, I will only focus on two groups of nintedanib 150 mg bid and placebo since a main purpose of the review is to confirm the success of nintedanib 150 mg bid in the key secondary endpoints specified in phase 3 studies in addition to the primary endpoint of FVC decline . The following graphs describe the FVC change from baseline over time in each individual patient. Majority of patients seem to experience decline in FVC although degree of decline appears slightly lower in nintedanib group. In group mean graphs, the slope of decline in FVC of nintedanib group is smaller than the slope of placebo group (Figures 14-16).

Figure 14. FVC trend over time in individuals randomized to placebo (Study 1199.30)



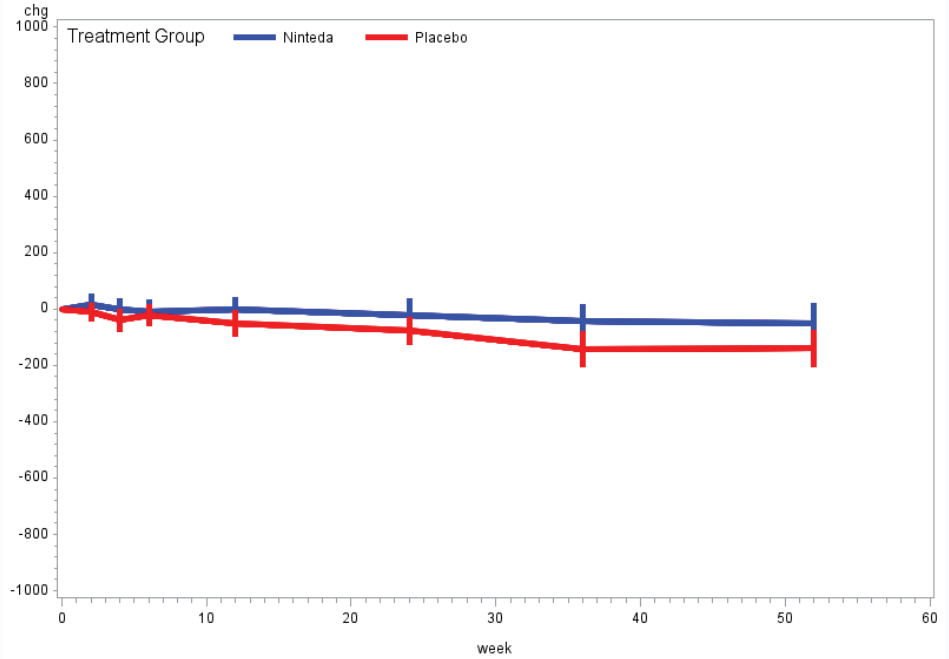
Source: Reviewer

Figure 15. FVC trend over time in individuals randomized to nintedanib (Study 1199.30)



Source: Reviewer

Figure 16. Mean FVC trend over time by treatment group (Study 1199.30)



Source: Reviewer

Patients receiving nintedanib had a smaller mean decline from Baseline in FVC compared to those receiving placebo at Week 52 (p =0.014, random coefficients regression model) (Table 12). An estimated absolute difference was 131 mL FVC between the two treatment groups. Sensitivity analyses with respect to statistical model and missing data by me gave results consistent with the primary analysis.

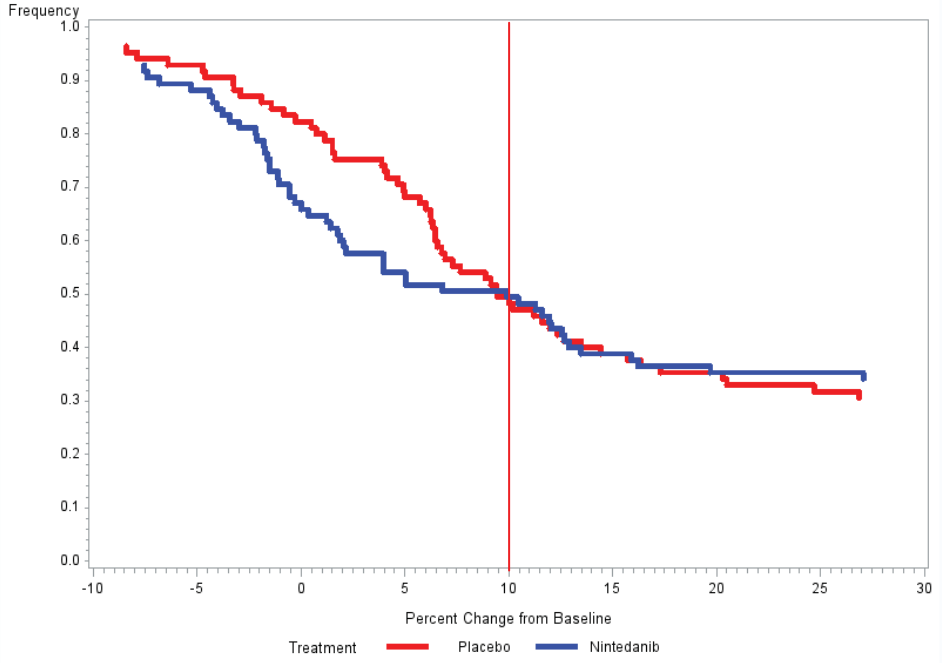
Table 12. Analyses on Annual Rate of Decline in FVC from Baseline to Week 52 (Study 1199.30)

Treatment	N	LS Mean	Difference vs. PBO		
			LS Mean Difference	95% CI	P-value
Applicant's Primary Analysis: ITT (Random Coefficient Regression*)					
Nintedanib 150mg bid	84	-60	131	(27, 235)	0.014
Placebo	83	-191			
My Primary Analysis: ITT (Random Coefficient Regression with same model used in Phase 3 studies)**					
Nintedanib	84	-75	140	(42, 238)	0.005
Placebo	83	-215			
My Sensitivity Analysis: ITT (ANCOVA with Placebo Mean Imputation)					
Nintedanib	85	-87	95	(17, 173)	0.018
Placebo	85	-182			
My Sensitivity Analysis: ITT (Rank ANCOVA with Lowest Rank Imputation)					
Nintedanib	85	-87	95	(17, 173)	0.004
Placebo	85	-182			

Source: Reviewer
Based on a MMRM with terms for treatment time, gender* height, gender* age, patient effect, patient* time (patient effect and patient* time random, all other effects fixed)
**Based on a MMRM with terms for treatment* time, gender, height, age, patient effect, patient* time (patient effect and patient* time random, all other effects fixed)

I also conducted a continuous responder analysis. The positive treatment effect of nintedanib was demonstrated by consistent separation of the curve across different level of response. Separation between curves of two groups was shown mostly in 10% or less relative decline (Figure 17).

Figure 17. Cumulative distribution of relative change from baseline in FVC (Study 1199.30)



Source: Reviewer

In summary, the phase 2 study in patients with IPF, showed statistically significant evidence in favor of nintedanib on the change in lung function (primary efficacy endpoint). Several secondary analyses were conducted on the primary efficacy endpoint to assess the robustness of the primary analysis.

Secondary Efficacy Endpoints

I was able to confirm the results of the applicant's analyses of the secondary endpoints. A review of the two pre-specified secondary efficacy endpoints is included as the phase 3 studies failed to provide replicated evidence with respect to time to first acute IPF exacerbation and SGRQ total score. I also present the analysis of all-cause mortality for this study and integrated with the phase 3 studies.

The Time to First Acute IPF Exacerbation

The applicant's results for time to first acute exacerbation analysis are summarized in Table 13 and Figure 18. Kaplan-Meier estimates were used to summarize time to first acute IPF exacerbation and treatment differences were analyzed using the log-rank test. The hazard ratio (HR) was determined based on the Cox proportional hazard model, adjusted for sex, baseline height, and baseline age. Censoring was applied at 372 days after randomization. Patients for whom no exacerbation event was reported within 373 days (included) of randomization were censored at 373 days or last contact date, whichever occurred first.

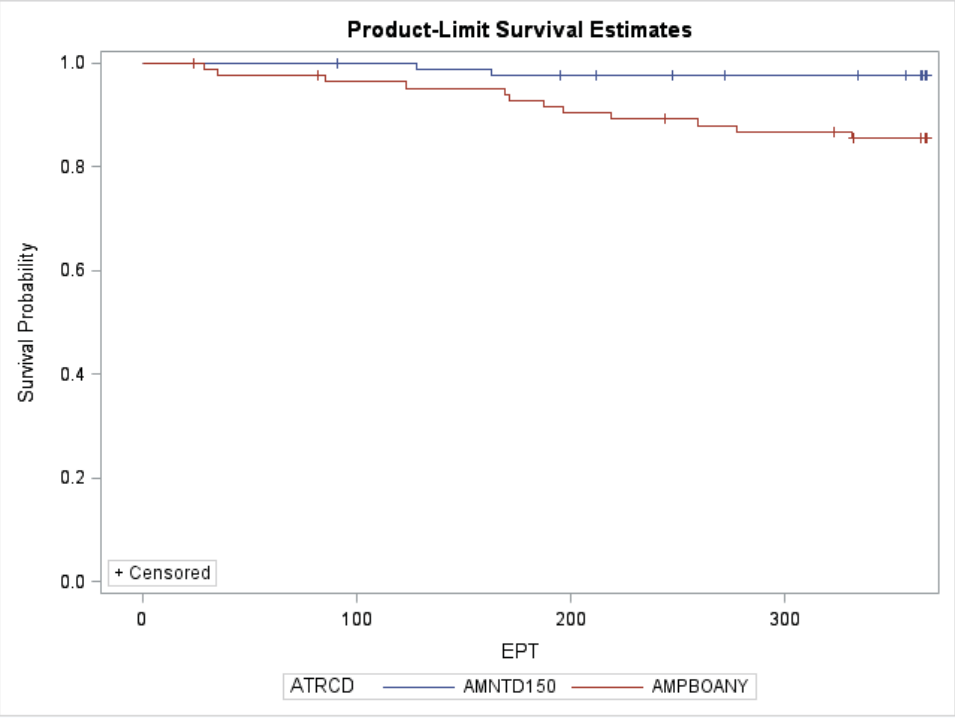
Treatment with nintedanib resulted in a lower proportion of at least one acute exacerbation than treatment with placebo (2.3%, 2/86 vs. 13.8%, 12/87 of patients, respectively). Treatment with nintedanib was associated with a 62% relative reduction of the risk of acute exacerbation compared to placebo (HR [95% CI]: 0.38 [0.19–0.77], p=0.016). There was also evidence of a treatment effect of nintedanib that began at approximately Week 12 and extended toward Week 52. Since replicated success in this endpoint is considered important in supporting the efficacy evidence from primary endpoint, the significance of the endpoint from this study consists of such replication.

Table 13. Survival Analysis on Time to First Acute Exacerbation over 52 weeks (Study 1199.30)

	Nintedanib	***Placebo***	***Hazard Ratio (95% CI)*** [b]
	N of Event (%)	***N of Event (%)***	***p-value*** [a]
Study 1199.30			
N of Randomized	86	87	
Acute exacerbation	2 (2.3)	12 (13.8)	0.16 (0.04, 0.71), 0.016

Source: Reviewer
[a] p-value was based on the log-rank test.
[b] Hazard ratio was based on the Cox proportional hazard model with terms for treatment, sex, age, and height.

Figure 18. Kaplan-Meier Curve of Time to First Acute Exacerbation over 52 weeks (Study 1199.30)



Source: Reviewer

The Change from baseline in SGRQ total score

The results from the analyses of the mean change from baseline in SGRQ are summarized in Table 14. The endpoint was analyzed using ANCOVA model with fixed effects for treatment, baseline SGRQ total score, and region after imputing missing data with LOCF method. The applicant also presented a sensitivity analysis using the same model after imputing missing data with the worst observed value carried forward (WOCF) method.

The mean change in SGRQ total score in patients treated with nintedanib is significantly lower compared to patients treated with placebo (-0.7 vs. 5.5, respectively; difference of -6.1, p=0.007). Applicant's sensitivity analysis and my sensitivity analyses with ANCOVA model after imputing missing data with the mean of placebo completers were consistent with results from the applicant's pre-specified analysis. Cumulative distribution curves by me crossed with worst score imputation for missing data, but showed separation of curves after -40% change or 40% worsening (Figure 15). Again since replicated success in this endpoint is considered important in supporting the efficacy evidence from primary endpoint, the significance of the endpoint from this study consists of such replication.

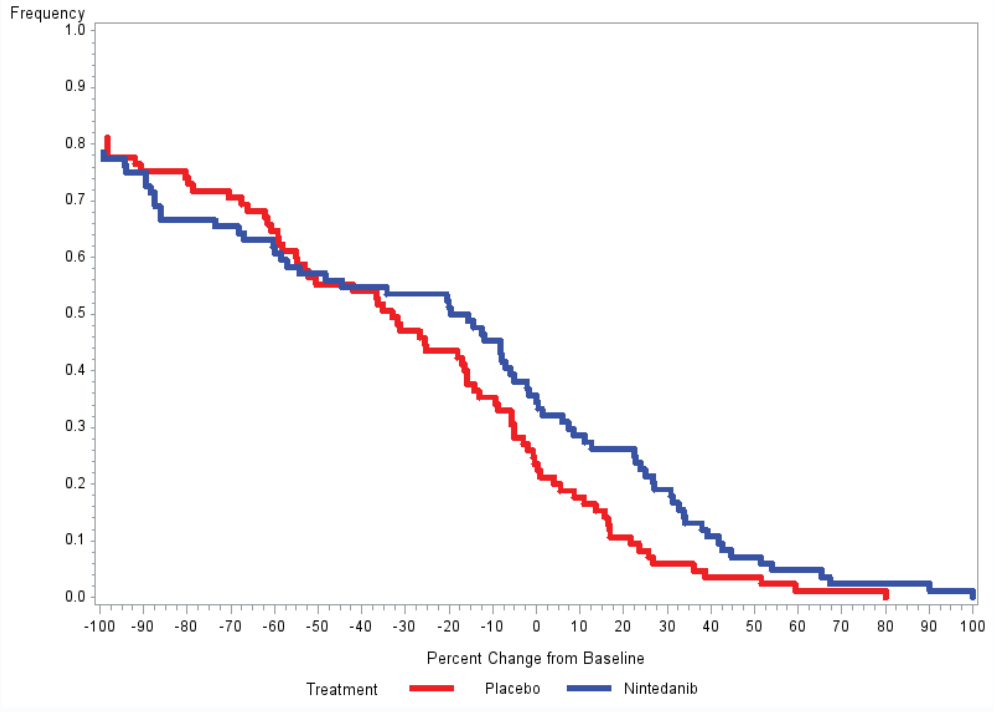
Table 14. Change from Baseline in SGRQ total score at Week 52 (Study 1199.30)

	Study 1199.30	
	Nintedanib (n=86)	Placebo (n=87)
Applicant's ANCOVA analysis with LOCF imputation		
N	75	79
LSMEAN (SE)	-0.7 (1.7)	5.5 (1.7)
vs. Placebo	-6.1 (2.3)	
95% CI	(-10.6, -1.7)	
p-value	0.007	
Applicant's ANCOVA analysis with WOCF imputation		
N	75	79
LSMEAN (SE)	0.4 (1.7)	6.0 (1.7)
vs. Placebo	-5.6 (2.2)	
95% CI	(-10.0, -1.2)	
p-value	0.013	
My ANCOVA analysis with placebo mean imputation		
N	86	87
LSMEAN (SE)	-1.0 (1.4)	3.9 (1.4)
vs. Placebo	-4.9 (1.9)	
95% CI	(-8.7, -1.1)	
p-value	0.012	

Source: Reviewer

Note: WOCF stands for worst-observed value-carried-forward.

Figure 19. Cumulative distribution of relative change from baseline in SGRQ total score (Study 1199.30)



Source: Reviewer

All-cause mortality

The applicant conducted an analysis comparing all-cause mortality between the treatment groups. Kaplan-Meier estimates were used to summarize survival time up to the end of the study treatment period. Survival time is measured by time from randomization to death. Treatment differences were analyzed using the log-rank test. The hazard ratio (HR) was determined based on the Cox proportional hazard model, with sex, age, and height as factors. The results are displayed in Table 15. Although there was numerically smaller proportion of deaths in nintedanib group compared to placebo group, the difference was not statistically significant (Figure 20).

Since this study was not powered for the survival endpoint, a post-hoc analysis pooling the three studies was conducted to increase the power to detect difference if any. However, the results on pooled data still did not reach a statistical significance (6%, 42/723 vs. 8%, 42/508; HR [95% CI]: 0.70 [0.46, 1.08], p=0.096) (Figure 21).

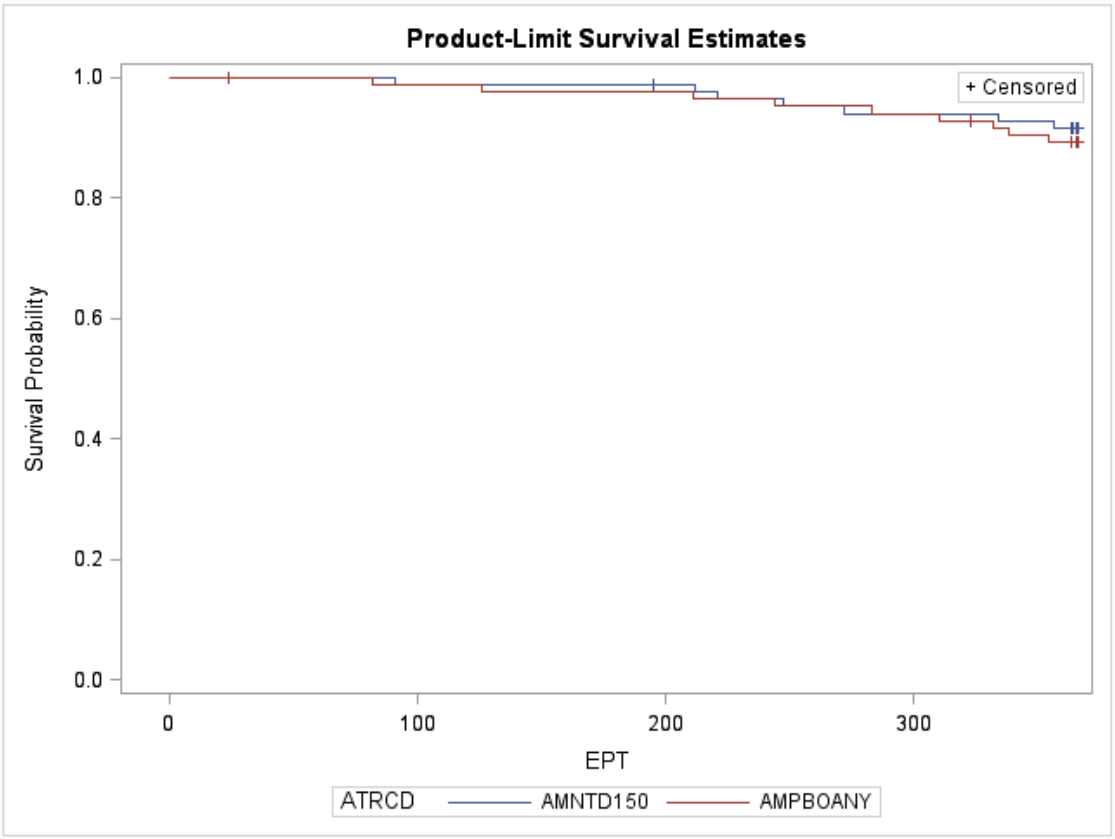
Table 15. Survival Analysis on All-Cause Mortality during the Treatment Period (All Treated Patients)

	Nintedanib	***Placebo***	***Hazard Ratio (95% CI)*** [c]
	N of Event (%)	***N of Event (%)***	***p-value*** [b]
Study 1199.30			
N of TS	86 [a]	87 [a]	
Death	7 (8.1)	9 (10.3)	0.73 (0.27, 1.98), 0.586
Censored	79 (91.9)	78 (89.7)	
Studies 1199.32, 1199.34, 1199.30 pooled [d]			
N of TS	723 [a]	508 [a]	0.70 (0.46, 1.08), 0.096
Death	42 (5.8)	42 (8.3)	
Censored	681 (94.2)	466 (91.7)	

Source: Reviewer

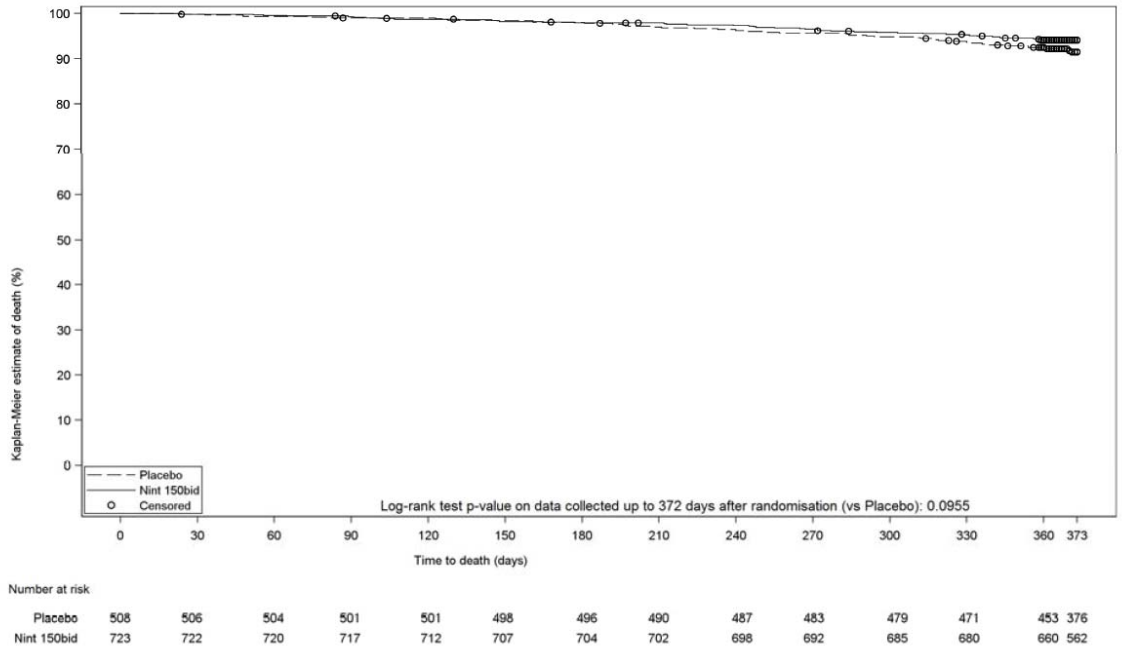
[a] Based on occurrence of event, or censoring in the absence of the event. Time to event was the event date minus randomization date plus one. The censoring date was the last available contact date or time of rescue (if one occurred), or the end of the Treatment Period.
[b] p-value was based on the log-rank test comparing nintedanib with placebo.
[c] Hazard ratio was based on the Cox proportional hazard model with terms for treatment, sex, age, and height.
[d] Hazard ratio was based on the Cox proportional hazard model with terms for study, treatment, sex, age, and height.

Figure 20. Kaplan-Meier Curve of Time to All-Cause Mortality during the Treatment Period (Study 1199.30)



Source: Reviewer

Figure 21. Kaplan-Meier Curve of Time to All-Cause Mortality during the Treatment Period (Studies 1199.32, 1199.34, and 1199.30 pooled)



Source: Excerpted from the response to Information Request #3 on Pooled Mortality Analysis (page 12).

3.3 Evaluation of Safety

The assessment of the safety of the study drug was mainly conducted by the reviewing medical team. The reader is referred to Dr. Miya Paterniti's review for information regarding the safety profile of the drug.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The following analyses are tabular and graphical presentation of the subgroup analyses by demographics, region, and baseline disease characteristics in terms of FVC change from baseline at Week 52. The subgroup analyses were consistent with the results from the overall population in terms of FVC change (Table 16 & Figure 22).

Table 16. Reviewer's Subgroup Analyses on FVC– Studies 1199.32 and 1199.34 pooled

	Nitedanib		Placebo		
	N	Mean	N	Mean	ABS Diff (95% CI)
Overall (p<0.001) [a]					
	638	-113	423	-222	109 (75, 143)
Sex (p=0.901) [b]					
Males	507	-118	334	-233	115 (75, 155)
Female	131	-92	89	-182	90 (32, 148)
Age (p=0.176) [b]					
<65 yrs	258	-126	145	-241	114 (55, 174)
≥65 yrs	380	-105	278	-210	106 (64, 147)
Region (p=0.541) [b]					
ROW [c]	540	-116	361	-218	102 (66, 139)
USA	98	-94	62	-241	147 (56, 238)
Race (p=0.896) [b]					
White	360	-109	248	-231	122 (78, 167)
N-White	278	-118	175	-209	90 (39, 142)
Baseline FVC (P=0.175) [b]					
<Median	323	-121	215	-212	92 (46, 137)
≥Median	315	-105	208	-231	127 (77, 177)
Smoke History (P=0.977) [b]					
Never S	174	-134	122	-223	89 (33, 146)
Smoke	464	-105	301	-222	117 (76, 159)
Time Since IPF Diagnosis (P=0.908) [b]					
<1 yrs	274	-109	193	-226	117 (63, 170)
≥1 yrs	364	-115	230	-219	104 (60, 148)

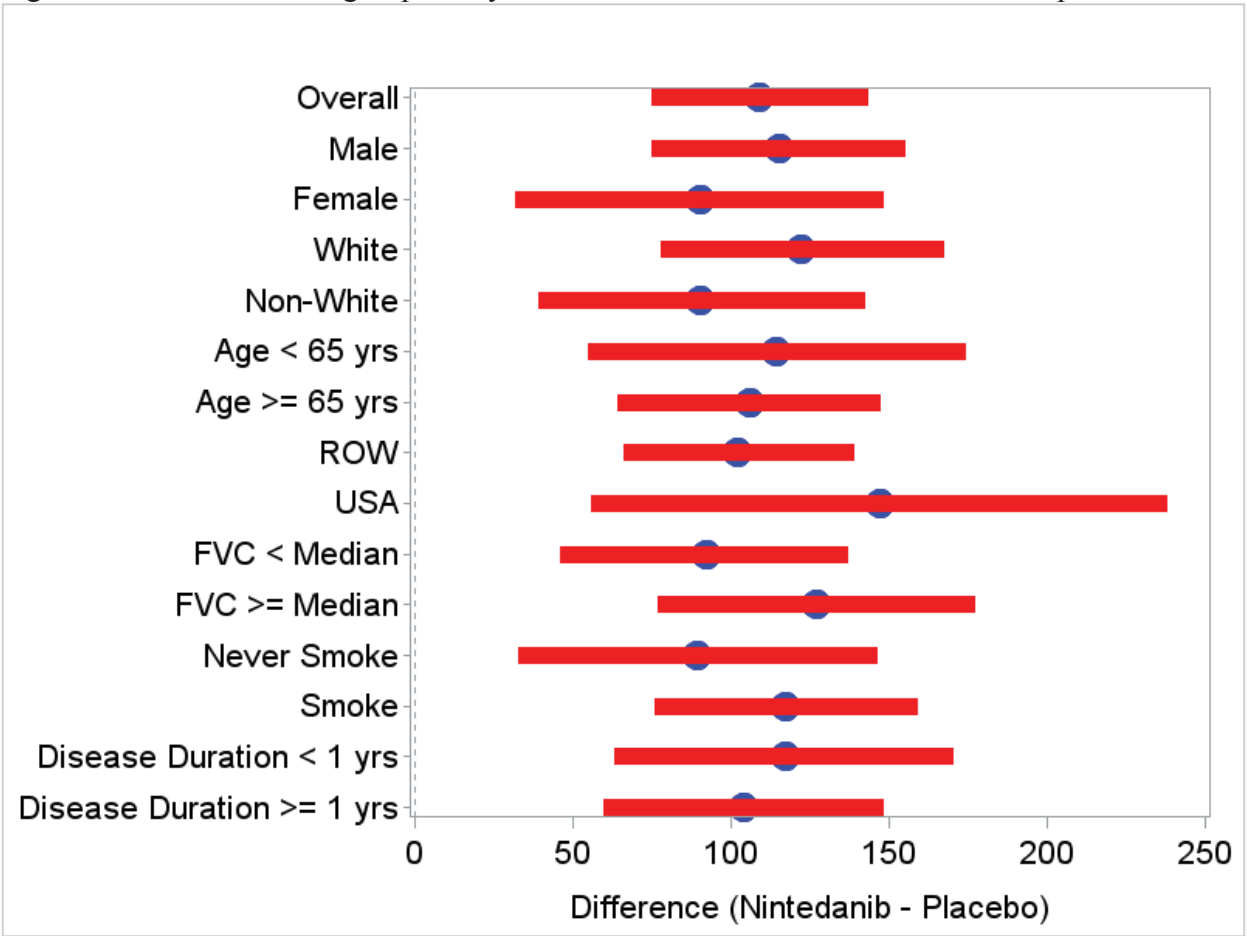
Source: Reviewer

[a] Random coefficiets regression model, comparing nintedanib to placebo.

[b] Random coefficiets regression model for interaction between treatment arm and subgroup.

[c] ROW includes Australia, Belgium, Canada, Chile, China, Czech Republic, Finland, France, Germany, Greece, India, Ireland, Israel, Italy, Japan, Korea, Mexico, Netherlands, Portugal, Russia, Spain, Turkey and the United Kingdom.

Figure 22. Reviewer's Subgroup Analyses on FVC– Studies 1199.32 and 1199.34 pooled



5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

During my review of the application, several potential statistical issues were identified, including the approach to handle missing data and multiplicity. There was also concern regarding substantial evidence of efficacy for several key secondary endpoints and a post hoc pooled analysis on all-cause mortality.

When analyzing the primary endpoint with random coefficients linear regression model, the applicant pre-specified to not impute missing data assuming a linear decline in lung function after treatment discontinuation. I find that this approach is acceptable since estimated slope based on available data from an individual patient before treatment discontinuation conservatively predicts the annual decline in lung function similar to linear extrapolation. The

applicant conducted a sensitivity analysis with multiple imputation based on various patterns of availability of data such as completers, retrieved dropouts, live dropouts, and dead dropouts. I think that the analyses with multiple imputations by the patterns provided useful information on the impact of missing data on the results. I also conducted a sensitivity analysis with imputation to missing data using the mean of placebo completers to penalize early dropouts in nintedanib group with good results before treatment discontinuation. In all cases, there was a significant treatment effect in favor of nintedanib.

In terms of multiplicity, the applicant proposed a hierarchical testing strategy for the key secondary endpoints. To test the secondary endpoints, the primary endpoint must be statistically significant at 5% level. Then they test on the secondary endpoints using sequential test procedure in the pre-specified order between the two key secondary endpoints, i.e., time to first acute IPF exacerbation and then change from baseline in SGRQ total score at 52 weeks.

Although the applicant won on the primary endpoint in two phase 3 studies, the key secondary endpoints, time to first acute IPF exacerbation and change from baseline in SGRQ total score, were only significant in one of the phase 3 studies, 1199.34. To provide replicated evidence for these secondary endpoints, I examined result from a phase 2 study, 1199.30. With this evidence from the phase 2 study, overall there seems substantial evidence of efficacy for both the primary and key secondary endpoints.

All-cause mortality was not shown to be statistically significantly different between nintedanib 150 mg bid and placebo although there was a trend favoring nintedanib. This was expected since the individual studies were not powered for the mortality endpoint. In order to increase power, mortality data from the two phase 3 studies and the phase 2study were integrated. However, the trend favoring nintedanib in the pooled analysis still did not reach the statistical significance.

Findings from the review of studies 1199.32, 1199.34, and 1199.30 are summarized below.

Primary Endpoint – Annual rate of decline in FVC (mL/yr)
Patients receiving nintedanib 150 mg bid had a lower mean annual rate of decline from Baseline in FVC compared to those receiving placebo at Week 52 ($p < 0.001$, random coefficients linear regression model) in both studies 1199.32 and 1199.34. This represents an absolute difference of 125 mL/year and 94 mL/year in rate of decline between the two treatment groups in Study 1199.32 and Study 1199.34, respectively.

Key Secondary Endpoints – Time to First Acute IPF Exacerbation
Overall, treatment with nintedanib 150 mg bid resulted in a lower proportion of acute IPF exacerbation than treatment with placebo (4%, 12/329 vs. 10%, 21/219 of patients, respectively) in Study 1199.34. Treatment with nintedanib 150 mg bid was associated with a 62% relative reduction of the risk of acute IPF exacerbation compared to placebo (HR [95% CI]: 0.38 [0.19–0.77], p=0.005). This finding was not replicated in Study 1199.32 (HR [95%

CI]: 1.15 [0.54–2.42], p=0.623). However, treatment with nintedanib 150 mg bid resulted in a lower proportion of acute exacerbation than treatment with placebo (2%, 2/86 vs. 14%, 12/87 of patients, respectively) in Study 1199.30. Treatment with nintedanib150 mg bid was associated with an 84% relative reduction of the risk of acute exacerbation compared to placebo (HR [95% CI]: 0.16 [0.04–0.71], p=0.005).

Key Secondary Endpoints – Change from Baseline in SGRQ total score
Patients receiving nintedanib 150 mg bid had a lower mean change from Baseline in SGRQ total score compared to those receiving placebo at Week 52 in Study 1199.34 (difference= -2.69, p =0.020, mixed effects repeated measures (MMRM) model). A statistically significant difference of change from baseline in SGRQ total score at Week 52 between the two treatment groups was not achieved in Study 1199.32 (difference= -0.05, p =0.966, MMRM model). However, a statistically significant difference of change from baseline in SGRQ total score at Week 52 between the nintedanib 150 mg bid and placebo was shown in Study 1199.30 (difference= -6.12, p =0.007, Analysis of Covariance (ANCOVA) model).

Other Endpoint – All-cause Mortality
Difference in all-cause mortality was not shown statistically significant in either phase 3 studies, 1199.32 and 1199.34 as expected since the studies were not powered for this rare event. However, there was some numerical evidence in favor of nintedanib in Study 1199.32 (4%, 13/309 vs. 6%, 13/204; HR [95% CI]: 0.63 [0.29, 1.36], p=0.288) and in Study 1199.34 (7%, 22/329 vs. 9%, 20/219; HR [95% CI]: 0.74 [0.40, 1.35], p=0.300). All-cause mortality over 52 weeks on pooled data from studies 1199.30, 1199.32, and 1199.34 were analyzed post-hoc by the applicant as response to Information Request. When data from three studies were pooled, the applicant stated there was some evidence of survival benefit in the nintedanib 150 mg bid group compared to placebo on all-cause mortality (6%, 42/723 vs. 8%, 42/508) although it still did not reach the statistical significance (HR [95% CI]: 0.70 [0.46–1.08], p=0.096).

5.2 Conclusions and Recommendations

The efficacy data from studies 1199.32, 1199.34, and 1199.30 provide substantial evidence of nintedanib 150 mg bid for treatment of IPF based on the annual rate of decrease in lung function by FVC and improvement in quality of life based on SGRQ and delay of acute IPF exacerbation. While three studies showed statistical significance on advantage of nintedanib on lung function based on FVC data, Study 1199.34 and Study 1199.30 showed improvement in quality of life based on SGRQ data and delay of acute IPF exacerbation based on time to event data.

5.3 Labeling Recommendations

Following is an excerpt from the relevant clinical studies section in the proposed label. I generally agree with the study description and primary analysis results and their interpretation.

However, we recommend that results from secondary and exploratory analyses that were not adjusted for multiplicity be not presented except for endpoints agreed as clinically important. All integrated analyses on efficacy data should be deleted except for the all-cause mortality data.

(b) (4)

[redacted]

4 Pages have been Withheld in Full as b4 (CCI/ TS) immediately following this page.

APPENDICES

Table 17. Proportion of FVC Responders at 52 Weeks (Study 1199.32)

	Placebo	Nintedanib 150 mg bid
Number of patients in TS [N (%)]	204 (100.00)	309 (100.00)
5% threshold		
Number of FVC responders (%)[1]	78 (38.24)	163 (52.75)
95% CI	(31.84, 45.06)	(47.18, 58.25)
Comparison vs placebo		
Odds ratio		1.847
95% CI		(1.28, 2.66)
p-value[2]		0.0010
10% threshold		
Number of FVC responders (%)[1]	116 (56.86)	218 (70.55)
95% CI	(50.00, 63.47)	(65.24, 75.36)
Comparison vs placebo		
Odds ratio		1.914
95% CI		(1.32, 2.79)
p-value[2]		0.0007

[1] Responder patients were those with no absolute decline greater than 5% or 10%, respectively, in FVC% predicted and in patients with an FVC evaluation at 52 weeks

[2] Based on logistic regression with terms treatment, age, gender, height and baseline FVC % predicted included Patients with missing data were considered as non-responders

Source: Excerpted from the Clinical Study Report of Study 1199.32 (page 125).

Table 18. Proportion of FVC Responders at 52 Weeks (Study 1199.34)

	Placebo	Nintedanib 150 mg bid
Number of patients in TS [N (%)]	219 (100.00)	329 (100.00)
5% threshold		
Number of FVC responders (%)[1]	86 (39.27)	175 (53.19)
95% CI	(33.04, 45.87)	(47.79, 58.52)
Comparison vs placebo		
Odds ratio		1.794
95% CI		(1.26, 2.55)
p-value[2]		0.0011
10% threshold		
Number of FVC responders (%)[1]	140 (63.93)	229 (69.60)
95% CI	(57.38, 70.00)	(64.43, 74.33)
Comparison vs placebo		
Odds ratio		1.286
95% CI		(0.89, 1.86)
p-value[2]		0.1833

[1] Responder patients were those with no absolute decline greater than 5% or 10%, respectively, in FVC% predicted and in patients with an FVC evaluation at 52 weeks

[2] Based on logistic regression with terms treatment, age, gender, height and baseline FVC %predicted included Patients with missing data were considered as non-responders

Source: Excerpted from the Clinical Study Report of Study 1199.34 (page 123).

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

YONGMAN KIM
09/02/2014

DAVID M PETULLO
09/03/2014

THOMAS J PERMUTT
09/03/2014
I concur.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CARCINOGENICITY STUDIES

NDA/BLA #:	NDA 205832/S-000 (reference: IND 74683)
Drug Name:	Nintedanib (BIBF 1120)
Indication(s):	Treatment for Idiopathic Pulmonary Fibrosis (IPF)
Applicant:	Boehringer Ingelheim Pharmaceuticals, Inc. (BI)
Date(s):	Received 5/2/2014
Documents Reviewed:	Study DDB0006 (Mice) and Study DDB0007 (Rats) reports and electronic datasets submitted with the electronic submission on 6/6/2014
Review Priority:	Priority Review
Biometrics Division:	Division of Biometrics VI
Statistical Reviewer:	Feng Zhou, M.S.
Concurring Reviewers:	Karl Lin, Ph.D.
Medical Division:	Division of Pulmonary, Allergy and Rheumatology Products
Clinical Team:	Carol M. Galvis, Ph.D.
Project Manager:	Jessica K. Lee
Keywords:	Carcinogenicity, Dose response

Table of Contents

1 Summary

This review evaluates statistically the tumorigenicity data of 2-year oral carcinogenicity studies of nintedanib (BIBF 1120 ES) in rats and mice. Nintedanib is tyrosine kinase (TK) inhibitor. The drug inhibits TKs associated with receptors for platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), and vascular endothelial growth factor (VEGF). The review analyzes the dose-response relationship of tumor incidence, mortality, and tumor-related mortality. The analysis reveals no evidence of nintedanib tumorigenicity from the statistical perspective.

Mouse Study: Mice (66/sex/dose) were treated with 0, 5, 15, or 30-mg/kg/day nintedanib for up to 104 weeks. The control group (0 mg/kg/day) received the vehicle only.

Survival analysis showed dose-related and statistically significant trend in decreases in survival rates in males ($p<0.001$), but not in females. Pair-wise comparison showed statistically significant increase in mortality in the high dose (HD) group in both sexes ($p<0.001$ and $p=0.026$ for male and female mice, respectively) when compared with the control group. In fact, the 30-mg/kg/day male group was terminated at study week 103 when the survival reached 23%. In the control and other treated male groups, between 55 and 58% of animals survived until termination in Week 103. The female animals were terminated at study week 104.

Tests on tumor data show a positive trend ($p = 0.027$) almost at the significance level of 0.025, and a statistically significant increase for mammary adenocarcino tumor in female mice. The trends in incidence in the combination of mammary adenoacantho, mammary adenocarcino and mammary adenoma in female mice and in the combination of malignantlymphoma and plasma cell lymphoma (per request by the pharmacologist Dr. Galvis) in female mice were not statistically significantly affected by treatment. (See the following table for details)

Tumor Incidence Rates in Female Mice

Organ Name	Tumor Name	0 mg Cont N=66	5 mg Low N=66	15 mg Med N=66	30 mg High N=66	P-Value Dos Resp	C vs. L	C vs. M	C vs. H
H-POIETIC TUMOU	HISTIOCYTIC SARCOMA	3	4	3	3	0.4858	0.4530	0.6415	0.5984
	MALIGNANT LYMPHOMA	**14**	**17**	**17**	**11**	**0.7138**	**0.3107**	**0.3343**	**0.5669**
	MYELO D CELL LEUKAEM	1	1	0	0	0.7958	0.7256	0.4845	0.4624
	PLASMA CELL LYMPHOMA	**0**	**0**	**0**	**1**	**0.2337**	**.**	**.**	**0.4674**
	COMB_LYMPHOMA	14	17	17	12	0.6245	0 3107	0.3343	0.4708
MAMMARY AREAS	ADENOMYOEPITHELIOMA	0	0	0	1	0 2337	.	.	0.4674
	MAMMARY ADENOACANTHO	**0**	**1**	**0**	**1**	**0.2886**	**0.4787**	**.**	**0.4674**
	MAMMARY ADENOCARCINO	**0**	**2**	**1**	**4**	**0.0270***	**0.2318**	**0.4948**	**0.0465***
	MAMMARY ADENOMA	**1**	**0**	**0**	**0**	**0.7337**	**0.4787**	**0.4896**	**0.4674**
	COMB_MAMMARY	1	3	1	5	0.0581	0 2849	0.7474	0.0791

Rat Study: Rats (60/sex/dose) were treated with 0, 2.5, 5, or 10-mg/kg/day nintedanib for up to 104 weeks. The control group (0 mg/kg/day) received the same volume of vehicle. Analyses for end points that include tumor incidence, survival, and tumor-related deaths did not reveal any evidence of tumorigenicity in either sex.

2 Background

The sponsor conducted a 24-month carcinogenicity study by oral (gavage) administration in male and female CD-1 mice and a 24-month carcinogenicity study by oral (gavage) administration in Han Wistar rats. This review analyzed the SAS datasets of these studies received on 06/06/2014 via submission NDA205832 (b) (4)

In this review these dose groups would be referred to as the low (L), medium (M), and high (H) dose groups, respectively. The phrase "dose response relationship" refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases. Results of this review have been discussed with the reviewing pharmacologist Dr. Galvis.

3 Rat Study

Study Report: DDB0007-final-report.pdf; SAS data: Tumor.xpt

In this study the carcinogenic potential of BIBF 1120 was assessed in Han Wistar rats. The test material was administered by oral (gavage) at doses of 2.5, 5.0 or 10.0 mg/kg/day. A similarly constituted control group received the vehicle, 0.5% hydroxyethyl cellulose (Natrosol® 250 HX) in demineralized water, at the same volume dosage. A further five males and five females were allocated to the control group, and 10 males and 10 females were allocated to each treated group and were used for toxicokinetic evaluation.

Two separate experiments were conducted, one in male and one in female rats. All males and females were terminated after 104 weeks. There were 275 males and 275 females assigned randomly to control and three treated groups of equal size in each main experiment.

Animal Group Assignments

Group	Treatment	Dose (mg/kg/day) #	Number of animals			
			Main study		Satellite study†	
			Male	Female	Male	Female
1	Control	0	60	60	5	5
2	BIBF 1120 ES	2.5	60	60	10	10
3	BIBF 1120 ES	5	60	60	10	10
4	BIBF 1120 ES	10	60	60	10	10

A conversion factor for salt/base ratio was used. 1.000 g of the base corresponds to 1.204 g of the salt form
† Satellite animals used for toxicokinetic sampling only

During the administration period all rats were observed twice daily for morbidity and mortality (once on weekends and public holidays). A detailed clinical examination was performed once before the start of treatment and once weekly thereafter. The rats were palpated regularly for the appearance of masses during the clinical observations.

Body weights of all rats were measured once before treatment commenced (week -1), on the day that treatment commenced (Week 0), at weekly intervals for the first 16 weeks of treatment, thereafter once every four weeks and before necropsy. The rats in the control group remained untreated.

3.1 Sponsor's Analyses

3.1.1 Survival Analysis

The sponsor performed survival analysis using the Kaplan-Meier product-limit method. An overall trend test comparing all groups was conducted using a log-rank test. If this overall trend test was significant ($p<0.05$), the highest dose group was excluded and the trend test repeated (using a one-tailed test), until the test was no longer statistically significant.

Sponsor's findings: Sponsor's analysis showed that mortality in all treatment groups was lower than in control groups. Statistical analysis of mortality when all male treated groups were included, was significant for trend ($p=0.003$). Upon exclusion of the 10 mg/kg/day treated group, the trend test was still significant ($p=0.030$) and upon further exclusion of the 5 mg/kg/day treated group, the trend test was no longer significant ($p=0.382$). The pairwise comparison of the control group with the 10 mg/kg/day treated group was statistically significant ($p=0.010$). For females the trend test was not statistically significant when all groups were included in the analysis ($p=0.151$) and none of the pairwise comparisons were statistically significant.

Survival at study termination for males was 39 (65%), 41 (68%), 49 (82%), and 52 (87%) in the control, low, median, and high dose group, respectively. Survival at study termination for females was 37 (62%), 46 (77%), 43 (72%), and 46 (77%).

The Sponsor's overall Survival for Male and Female Rats

	Group and sex							
	1M	2M	3M	4M	1F	2F	3F	4F
Dose (mg/kg/day)	0	2.5	5	10	0	2.5	5	10
Group size	60	60	60	60	60	60	60	60
No. of deaths								
Main study	21	19	11	8	23	14	17	14
Number of surviving males and females reaching termination								
Main study	39	41	49	52	37	46	43	46

[Source: page 43 of study report (ddb0007-final-report.pdf)]

3.1.2 Tumor Data Analysis

The sponsor analyzed the tumor incidence data using the methods outlined in the paper of Peto et al. (1980) for positive dose response relationships and the Fisher exact test for pairwise comparisons of the treated groups with the control. For Peto analysis the sponsor first classified the tumor types as fatal and incidental, and analyzed them using the death rate and prevalence methods, respectively. For the evaluation of incidental tumors, the experimental period was divided into partitions using the ad hoc run procedure described in Peto et al.

Adjustment for multiple testing**:** In order to control the false positive error, the sponsor tested the common and the rare tumors at 0.005 and 0.025 significance levels, respectively (Lin, 2000) for positive dose response relationship, and 0.01 and 0.05 for pairwise comparisons. Tumors are considered as common with a background rate of ≥ 1% and as rare with a background incidence of < 1%.

Sponsor's findings: The sponsor's analyses show that there were no statistically significant differences between the control group and the treated groups in both sexes.

3.2 Reviewer's Analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing pharmacologist, this reviewer independently performed survival and tumor data analyses. Data used in this reviewer's analyses were provided by the sponsor electronically in submission NDA 205832 (b) on 6/6/2014.

3.2.1 Survival Analysis

The survival distributions of rats in all treatment groups were estimated using the Kaplan-Meier product limit method. For control, low, medium, and high dose groups, the dose response relationship was tested using the likelihood ratio test and the homogeneity of survival distributions was tested using the log-rank test. The Kaplan-Meier curves for survival rates are given in Figures 1A and 1B in the appendix for male and female rats, respectively. The intercurrent mortality data are given in Tables 1A and 1B in the appendix for male and female rats, respectively. Results of the tests for dose response relationship and homogeneity of survivals, are given in Tables 3A and 3B in the appendix for male and female rats, respectively.

Reviewer's findings: This reviewer's analysis showed 21 (35%), 19 (32%), 11 (18%), and 8 (13%) number (percent) of deaths in male rats and 23 (38%), 14 (23%), 17 (28%), and 14 (23%) number (percent) of deaths in female rats in control, low, medium, and high dose groups, respectively. The tests showed statistically significant dose response relationship in mortality across control and treated groups in male rats. The pairwise comparisons showed statistically significant increased mortality in the male rat high dose and medium dose groups when compared with the control group and decreased mortality in the female rat high dose group compared to their respective control.

3.2.2 Tumor Data Analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups. Both the dose response relationship tests and pairwise comparisons were performed using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993). In this method an animal that lives the full study period ($w_{\max}$) or dies before the terminal sacrifice but develops the tumor type being tested gets a score of $s_h = 1$. An animal that dies at week w_h without a tumor before the end of the study gets a score of $s_h = \left(\frac{w_h}{w_{\max}}\right)^k < 1$.

The adjusted group size is defined as Σs_h. As an interpretation, an animal with score $s_h = 1$ can be considered as a whole animal while an animal with score $s_h < 1$ can be considered as a partial animal.

The adjusted group size Σs_h is equal to N (the original group size) if all animals live up to the end of the study or if each animal that dies before the terminal sacrifice develops at least one tumor, otherwise the adjusted group size is less than N. These adjusted group sizes are then used for the dose response relationship (or the pairwise) tests using the Cochran-Armitage test. One critical point for Poly-k test is the choice of the appropriate value of k, which depends on the tumor incidence pattern with the increased dose. For long term 104 week standard rat and mouse studies, a value of k=3 is suggested in the literature. Hence, this reviewer used k=3 for the analysis of this data. For the calculation of p-values the exact permutation method was used. The tumor rates and the p-values of the tested tumor types are listed in Tables 5A and 5B in the appendix for male and female rats, respectively.

Multiple testing adjustment: For the adjustment of multiple testing of dose response relationship, the FDA guidance for the carcinogenicity study design and data analysis suggests the use of test levels α=0.005 for common tumors and α=0.025 for rare tumors for a submission with two species, and a

significance level α=0.01 for common tumors and α=0.05 for rare tumors for a submission with one species in order to keep the false-positive rate at the nominal level of approximately 10%. A rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. For multiple pairwise comparisons of treated group with control the FDA guidance the suggested the use of test levels α=0.01 for common tumors and α=0.05 for rare tumors, in order to keep the false-positive rate at the nominal level of approximately 10% for both submissions with two or one species.

It should be noted that the FDA guidance for multiple testing for dose response relationship is based on a publication by Lin and Rahman (1998). In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Lin and Rahman (2008) showed that this rule for multiple testing for dose response relationship is also suitable for Poly-K tests.

Reviewer's findings: There is no tumor types showed p-values less than or equal to 0.05 either for dose response relationship or pairwise comparisons of treated groups and control.

4 Mouse Study

Study Report: DDB0006.pdf; SAS data: Tumor.xpt

In this study the carcinogenic potential of nintedanib was assessed in CD-1 mice. The test material was administered by oral (gavage) at doses of 5, 15 and 30 mg/kg/day. A similarly constituted control group received the vehicle, 0.5% hydroxyethyl cellulose (Natrosol® 250 HX) in demineralised water, at the same volume-dosage (10 mL/kg). A further nine males and nine females were allocated to the control group, and 18 males and 18 females were allocated to each treated group and were used to provide blood samples for toxicokinetic evaluation.

Two separate experiments were conducted, one in male and one in female mice. Males were terminated after 102 weeks of treatment, when the number remaining in the high dose group (30 mg/kg/day) reached 15. Mortality during the first six months was high in males receiving high dose, however, sufficient main study animals survived at least 102 weeks of treatment to allow a meaningful evaluation of carcinogenicity. Females were terminated after 104 weeks. There were 327 males and 327 females assigned randomly to control and three treated groups of equal size in each main experiment.

Animal Group Assignments

Group	Treatment	Dosage (mg/kg/day) #	Number of animals			
			Main study		Satellite study†	
			Male	Female	Male	Female
1	Control	0	66	66	9	9
2	BIBF 1120 ES	5	66	66	18	18
3	BIBF 1120 ES	15	66	66	18	18
4	BIBF 1120 ES	30	66	66	18	18

#All doses are expressed in terms of BIBF 1120 BS, the free base equivalent of BIBF 1120 ES
† Satellite animals used for toxicokinetic sampling only

A viability check was performed near the start and end of each working day. Animals were observed at least twice daily (in the morning and afternoon) for any signs of moribundity. Cages were inspected daily for evidence of animal ill-health amongst the occupants.

The weight of each mouse was recorded one week before treatment commenced (Week -1), on the day that treatment commenced (Week 0), at weekly intervals for the first 16 weeks of treatment, thereafter once every four weeks and before necropsy. The mice in the control group remained untreated.

4.1 Sponsor's Analyses

4.1.1 Survival Analysis

The sponsor performed survival analysis using the same methods that were used in the rat study.

Sponsor's findings: Sponsor's analysis showed that mortality was higher than that of the controls and other treated groups throughout the treatment period for males receiving 30 mg/kg/day, necessitating premature termination of this sex in Week 103 of treatment when survival in the high dose group fell to 23% (15 survivors). In the control and other treated male groups, between 55% and 58% of animals survived until termination in Week 103. The trend test, when all treated groups were included, was statistically significant ($p<0.001$). Upon exclusion of the 30 mg/kg/day treated group, the trend test was no longer significant ($p=0.522$). The pairwise comparison of the control group with the 30 mg/kg/day treated group was statistically significant ($p<0.001$).

For females, the trend test was not statistically significant when all groups were included in the analysis ($p=0.076$). The pairwise comparison of the control group with the 30 mg/kg/day treated group was statistically significant ($p=0.034$).

The sponsor claimed that the cause of death for three high dose males (#247, #267 and #276) were considered to be associated with aspiration of dose as indicated by the necropsy findings of aerated fluid in the trachea and/or firm lungs. These three deaths occurred between Weeks 6 and 12 and there were no further deaths associated with the dosing procedure after Week 12.

Survival at study termination for males was 55%, 58%, 56%, and 23% in the control, low, median, and high dose group, respectively. Survival at study termination for females was 50%, 33%, 38%, and 30%. Sponsor concluded that the highest dose of 30 mg/kg/day was associated with increased mortality, particularly in males, indicated that the maximum tolerated dose had been exceeded.

The Sponsor's overall Survival for Male and Female Mice

	Group and sex							
	1M	2M	3M	4M	1F	2F	3F	4F
Dose (mg/kg/day)	0	5	15	30	0	5	15	30
Group size (Main/Satellites)	66/9	66/18	66/18	66/18	66/9	66/18	66/18	66/18
No. of deaths								
Main study	30	28	29	51	33	44	41	46
Satellite study	4	12	9	14	5	6	10	14
Cause of death / reason for premature euthanasia (main study)								
Aspiration of dose	0	0	0	3	0	0	0	0
Skin abrasions	11	9	7	12	0	4	4	4
Other	19	19	22	36	33	40	37	42
Number of surviving Main study animals reaching termination								
Main study	36	38	37	15	33	22	25	20
% Survival	55%	58%	56%	23%	50%	33%	38%	30%

[Source: page 40 of study report (ddb0006.pdf)]

4.1.2 Tumor Data Analysis

The sponsor analyzed the tumor incidence data using the same method that was used in the rat study.

Sponsor's findings: The sponsor's analyses did not show a statistically significant dose response relationship among the treatment groups in any of the observed tumor type. The sponsor concluded that there was no evidence of carcinogenic potential. Histopathological changes attributed to

treatment were confined to the gall bladder of animals given 15 or 30 mg/kg/day and comprised increased incidences of ulceration, reactive epithelial hyperplasia and fibrosis.

4.2 Reviewer's Analyses

4.2.1 Survival Analysis

The Kaplan-Meier curves for survival rates of all treatment groups are given in Figures 2A and 2B in the appendix for male and female mice, respectively. The intercurrent mortality data of all treatment groups are given in Tables 2A and 2B in the appendix for male and female mice, respectively. Results of the tests for dose response relationship and homogeneity of survivals for control, low, medium, and high dose groups are given in Tables 4A and 4B in the appendix for male and female mice, respectively.

Reviewer's findings: This reviewer's analysis showed 29 (44%), 28 (42%), 29 (44%), and 50 (76%) number (percent) of deaths in male mice, and 32 (48%), 44 (67%), 41 (62%), and 46 (70%) number (percent) of deaths in female mice in control, low, medium, and high dose groups, respectively. The tests showed a statistically significant dose response relationship in mortality across the treatment groups in male mice. The high dose mice showed a statistically significant increases in mortality in both sexes ($p<0.001$ and $p=0.026$ for male and female mice, respectively).

4.2.2 Tumor Data Analysis

The tumor data were analyzed for dose response relationships and pairwise comparisons of control group with each of the treated groups using the same method that was used for the rats study. The tumor rates and the p-values of the tested tumor types are listed in Tables 6A and 6B in the appendix for male and female rats, respectively.

Reviewer's findings: Following tumor types showed p-values less than or equal to 0.05 either for dose response relationship or pairwise comparisons of treated groups and control.

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups and Control in Male Mice

Organ Name	Tumor Name	0 mg Cont N=66	5 mg Low N=66	15 mg Med N=66	30 mg High N=66	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
LUNGS + BRONCHI	BRONCHIOLOALVEOLAR A	13	14	22	8	0 3980	0.4770	0.0419	0.4663

Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or Pairwise Comparisons of Treated Groups and Control in Female Mice

Organ Name	Tumor Name	0 mg Cont N=66	5 mg Low N=66	15 mg Med N=66	30 mg High N=66	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
MAMMARY AREAS	MAMMARY ADENOCARCINO	0	2	1	4	0 0270*	0 2318	0.4948	0.0465*

Based on the criteria of adjustment for multiple testing discussed above, none of the observed tumors was considered to have a statistically significant dose response relationship or a statistically significant increase in incidence in each treated group over the control group in male mice. Tests on tumor data of female mice show a positive trend ($p = 0.027$) almost at the significance level of 0.025, and a statistically significant increase in incidence in the high dose group over the control group for mammary adenocarcino tumor in female mice.

Dr. Galvis, the pharmacologist, requested for additional statistical analysis by combining some tumor types in female mice. The tumor types combined and their incidence rates and the results of statistical tests are presented in the table in the summary section at the beginning of this report. The results also

did not show any statistically significant dose response relationship and any statistically significant increase in incidence in individual treated groups over the control group in any of the observed combined tumor types (see also Table 6B in appendix for details).

5 Evaluation of the Validity of Design of Rat Study

As having been noted, except for the incidence of few tumors in female mice, no other tumor types showed statistically significant dose response relationship or increased incidence compared to their respective control. However, before drawing any conclusion regarding the carcinogenic or non-carcinogenic potential of the study drug in rats, it is important to look into the following two issues, as have been pointed out in the paper by Haseman (1984).

(i) Were enough animals exposed, for a sustained amount of time, to the risk of late developing tumors?
(ii) Were dose levels high enough to pose a reasonable tumor challenge to the animals?

There is no consensus among experts regarding the number of animals and length of time at risk, although most carcinogenicity studies are designed to run for two years with about fifty to sixty animals per treatment group. The following are some rules of thumb regarding these two issues as suggested by experts in this field.

Haseman (1985) has done an investigation on the first issue. He gathered data from 21 studies using Fischer 344 rats and B6C3Fl mice conducted at the National Toxicology Program (NTP). It was found that, on the average, approximately 50% of the animals in the high dose group survived the two-year study period. Also, in a personal communication with Dr. Karl Lin of Division of Biometrics-6, Haseman suggested that, as a rule of thumb, a 50% survival of 50 initial animals or 20 to 30 animals still alive in the high dose group, between weeks 80-90, would be consider as a sufficient number and adequate exposure. In addition Chu, Cueto and Ward (1981), suggested that "to be considered adequate, an experiment that has not shown a chemical to be carcinogenic should have groups of animals with greater than 50% survival at one-year."

It appears, from these three sources that the proportions of survival at 52 weeks, 80-90 weeks, and two years are of interest in determining the adequacy of exposure and number of animals at risk.

Regarding the question of adequate dose levels, it is generally accepted that the high dose should be close to the maximum tolerated dose (MTD). In the paper of Chu, Cueto and Ward (1981), the following criteria are mentioned for dose adequacy. A high dose is considered as close to MTD if any of the criteria is met.

I. "A dose is considered adequate if there is a detectable loss in weight gain of up to 10% in a dosed group relative to the controls."
II. "The administered dose is also considered an MTD if dosed animals exhibit clinical signs or severe histopathologic toxic effects attributed to the chemical."
III. "In addition, doses are considered adequate if the dosed animals show a slight increased mortality compared to the controls."

We will now investigate the validity of the riocignat rat and mouse carcinogenicity study, in the light of the above guidelines.

5.1 Rat Study

The following is the summary of survival data of rats in the high dose groups:

Percentage of Survival in the High Dose Group at the End of Weeks 52, 78, 91, and 102 in Rats

	Percentage of Survival			
Week	End of 52 Weeks	End of 78 Weeks	End of 91 Weeks	End of 104 Weeks
Male	97%	95%	93%	87%
Female	95%	93%	82%	77%

Based on the survival criterion Haseman proposed, it may be concluded that enough rats were exposed to the high dose for a sufficient amount of time in both sexes.

The following table shows the percent difference in mean body weight gain in rats from the concurrent control, defined as

$$\text{Percent difference} = \frac{(\text{Final BW} - \text{Baseline BW})_{\text{Treated}} - (\text{Final BW} - \text{Baseline BW})_{\text{Control}}}{(\text{Final BW} - \text{Baseline BW})_{\text{Control}}} \times 100$$

Percent Difference in Mean body Weight Gain from Controls in Rats

Male			Female		
Low	Medium	High	Low	Medium	High
1.22	4.16	-3.3	-3.1	-1 9	-5.0

Source: "Table 4 - Body weights – group mean values (g)" of Sponsor's report (Page 99)

Therefore, relative to the control the male rats in high dose group had about 3% and the female rats had about 5% decrements in their body weight gains.

The mortality rates at the end of the experiment were as follows:

Mortality Rates at the End of the Experiment in Rats

	Control	Low	Medium	High
Male	35%	32%	18%	13%
Female	38%	23%	28%	23%

This shows that the morality rates in the male rats high dose group is 22% lower than their control, while that in female rats is 15% lower than their control.

Thus, from the mortality and the body weight gain data it cannot be concluded that the high dose has reached the MTD in both sexes. For a final determination of the adequacy of the doses used, other clinical signs and histopathological toxic effects must be considered.

5.2 Mouse Study

The following is the summary of survival data of rats in the high dose groups:

Percentage of Survival in the High Dose Group at the End of Weeks 52, 78, 91, and 104 in Mice

	Percentage of Survival			
Week	End of 52 Weeks	End of 78 Weeks	End of 91 Weeks	End of 104 Weeks
Male	76%	52%	38%	24%
Female	95%	73%	50%	30%

Based on the survival criterion Haseman proposed, it may be concluded that enough mice were exposed to the high dose for a sufficient amount of time in both sexes.

The following table shows the percent difference in mean body weight gain in rats from the concurrent control based on the same definition as for rats study.

Percent Difference in Mean body Weight Gain from Controls in Mice

Male			Female		
Low	Medium	High	Low	Medium	High
3.70	1 39	-8.80	-17.35	-15.98	-30.59

Source: "Table 4 - Body weights – group mean values (g)" of Sponsor's report (Page 96)

Therefore, relative to the control the male rats in high dose group had about 9% and the female rats had about 31% decrements in their body weight gains.

The mortality rates at the end of the experiment were as follows:

Mortality Rates at the End of the Experiment in Rats

	Control	Low	Medium	High
Male	44%	42%	44%	76%
Female	48%	67%	62%	70%

This shows that the morality rates in the male rats high dose group is 22% higher than their control, while that in female rats is 12% higher than their control.

Thus, from the mortality and the body weight gain data it can be concluded that the used high dose level might have reached the MTD in male mice, and exceeded the MTD in female mice. For a final determination of the adequacy of the doses used, other clinical signs and histopathological toxic effects must be considered.

6 Conclusion

The 2-year oral carcinogenicity studies of nintedanib in rats and mice revealed no evidence of tumorigenicity in either species. Rats and mice (60 – 66/sex/dose) were treated with nintedanib daily by oral gavage for up to 104 weeks. The respective nintedanib dose in the control, low dose, mid dose and high dose groups was 0, 2.5, 5, and 10 mg/kg/day in rats and 0, 5, 15 and 30 mg/kg/day in mice. The high dose mice showed statistically significant increases in mortality in both sexes ($p<0.001$ and $p=0.026$ for male and female mice, respectively). The male mice treated with low and mid dose showed numerically significant, but statistically non-significant increases in mortality. There were no treatment-related in increases in tumor incidences in either sex in rats or mice. The review concludes that these 2-year carcinogenicity studies of nintedanib showed no evidence of tumorigenicity from the statistical perspective.

Feng Zhou, M.S.
Mathematical Statistician

Concur: Karl Lin, Ph.D.
Team Leader, Biometrics-6

cc:
Dr. Carol M. Galvis
Dr. Luqi Pei
Dr. Marcie Wood
Ms. Jessica K. Lee
Dr. Tsong
Ms. Zhou
Dr. Lin
Ms. Patrician

File Name: NDA205832Carcin.doc

7 Appendix

Table 1A: Intercurrent Mortality – Male Rats

	0 mg kg day (n=60)		2 mg kg day (n=60)		5 mg kg day (n=60)		10 mg kg day (n=60)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	0	0	1	1.67	1	1.67	2	3 33
53 - 78	3	5.00	3	6.67	3	6.67	1	5 00
79 - 91	11	23 33	6	16 67	4	13 33	1	6 67
92 - 104	7	35 00	9	31 67	3	18 33	4	13.33
Ter. Sac.	39	65 00	41	68 33	49	81 67	52	86.67

[#] Cum. %: Cumulative percentage except for Ter. Sac.

Table 1B: Intercurrent Mortality - Female Rats

	0 mg kg day (n=60)		2 mg kg day (n=60)		5 mg kg day (n=60)		10 mg kg day (n=60)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	1	1.67	.	.	.	.	3	5.00
53 - 78	6	11 67	3	5.00	5	8.33	1	6.67
79 - 91	8	25 00	5	13 33	6	18.33	3	11 67
92 - 104	8	38 33	6	23 33	6	28.33	7	23 33
Ter. Sac.	37	61 67	46	76 67	43	71.67	46	76 67

[#] Cum. %: Cumulative percentage except for Ter. Sac.

Table 2A: Intercurrent Mortality - Male Mice

	0 mg kg day (n=66)		5 mg kg day (n=66)		15 mg kg day (n=66)		30 mg kg day (n=66)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	5	7.58	5	7 58	4	6.06	16	24.24
53 - 78	7	18.18	8	19.70	9	19.70	16	48.48
79 - 91	5	25.76	9	33.33	6	28.79	9	62.12
92 - 102	12	43.94	6	42.42	10	43.94	9	75.76
Ter. Sac.	37	56.06	38	57.58	37	56.06	16	24.24

[#] Cum. %: Cumulative percentage except for Ter. Sac.

Table 2B: Intercurrent Mortality – Female Mice

	0 mg kg day (n=66)		5 mg kg day (n=66)		15 mg kg day (n=66)		30 mg kg day (n=66)	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	7	10.61	5	7.58	4	6.06	3	4 55
53 - 78	7	21.21	10	22.73	9	19.70	15	27.27
79 - 91	8	33.33	10	37.88	10	34.85	15	50.00
92 - 104	10	48.48	19	66.67	18	62.12	13	69.70
Ter. Sac.	34	51.52	22	33.33	25	37.88	20	30.30

[#] Cum. %: Cumulative percentage except for Ter. Sac.

Table 3A: Intercurrent Mortality Comparison – Male Rats

Test	Statistic	P-Value
Dose-Response	Likelihood Ratio	0 0016
Homogeneity	Log-Rank	0 0175

Table 3B: Intercurrent Mortality Comparison – Female Rats

Test	Statistic	P-Value
Dose-Response	Likelihood Ratio	0.1615
Homogeneity	Log-Rank	0.1851

Table 4A: Intercurrent Mortality Comparison – Male Mice

Test	Statistic	P-Value
Dose-Response	Likelihood Ratio	<0.0001
Homogeneity	Log-Rank	<0.0001

Table 4B: Intercurrent Mortality Comparison – Female Mice

Test	Statistic	P-Value
Dose-Response	Likelihood Ratio	0.0712
Homogeneity	Log-Rank	0.1184

Table 5A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Male Rats

Organ Name	Tumor Name	0 mg Control N=60	2.5 mg Low N=60	5 mg Med N=60	10 mg High N=60	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
ADIPOSE TISSUE	LIPOMA	0	0	0	1	0.2605	.	.	0.5234
ADRENALS	CORTICAL ADENOMA	1	0	0	0	0.7628	0 5096	0.5189	0.5234
	GANGLIONEUROMA	1	1	0	1	0.4734	0 2524	0.5140	0.2665
	L POMA	0	0	0	1	0.2605	.	.	0.5234
	PHAEOCHROMOCYTOMA	1	3	2	0	0.8526	0 3161	0.5212	0.5185
BONE	OSTEOMA	0	0	1	0	0.5163	.	0.5189	.
BRAIN	ASTROCYTOMA	1	1	0	0	0.8271	0 2573	0.5189	0.5234
	GRANULAR CELL TUMOUR	0	1	1	2	0.1368	0 5096	0.5189	0.2762
	PINEALOMA, MALIGNANT	0	1	0	0	0.5163	0 5096	.	.
DUODENUM	ADENOCARC NOMA	1	0	0	0	0.7593	0 5048	0.5140	0.5185
EPID DYM DES	MESOTHELIOMA	1	0	0	1	0.5185	0 5096	0.5189	0.2716
H-POIETIC TUMOU	LEUKAEMIA	1	0	0	0	0.7593	0 5048	0.5140	0.5185
	MALIGNANT LYMPHOMA	2	0	0	1	0.6186	0.7571	0.7662	0.5347
HEART	ENDOCARDIAL SCHWANNO	1	0	0	0	0.7593	0 5048	0.5140	0.5185
JEJUNUM	ADENOMA	0	0	1	0	0.5163	.	0.5189	.
	FIBROSARCOMA	1	0	0	0	0.7593	0 5048	0.5140	0.5185
	LEIOMYOMA	1	0	0	0	0.7628	0 5096	0.5189	0.5234
KIDNEYS	RENAL LIPOSARCOMA	0	0	0	1	0.2605	.	.	0.5234
LIVER	HEPATOCELLULAR ADENO	1	0	0	0	0.7628	0.5096	0.5189	0.5234
LN MESENTERIC	HAEMANGIOMA	1	2	4	0	0.7113	0 5146	0.2060	0.5234
	HAEMANGIOSARCOMA	1	0	0	0	0.7628	0 5096	0.5189	0.5234
LUNGS + BRONCHI	BRONCHIOLOALVEOLAR A	1	0	0	0	0.7628	0.5096	0.5189	0.5234
PANCREAS	AC NAR CELL ADENOCAR	0	0	1	0	0.5163	.	0.5189	.
	AC NAR CELL ADENOMA	1	0	0	0	0.7628	0 5096	0.5189	0.5234
	ISLET CELL ADENOMA	2	4	5	3	0.4541	0 3483	0.2425	0.5354
	ISLET CELL CARC NOMA	1	0	0	0	0.7593	0 5048	0.5140	0.5185
PARATHYROIDS	CHIEF CELL ADENOMA	0	1	0	2	0.1188	0 5096	.	0.2716
PITUITARY	ADENOMA, PARS DISTA	19	22	22	16	0.7933	0 2960	0.2960	0.6575
	ADENOMA, PARS INTERM	1	1	0	1	0.4734	0 2524	0.5140	0.2665
PREPUTIAL GLAND	SQUAMOUS CELL CARC N	0	0	1	0	0.5163	.	0.5189	.
	SQUAMOUS CELL PAPILL	2	0	0	0	0.9446	0.7619	0.7709	0.7752
PROSTATE	ADENOMA	0	1	0	0	0.5163	0 5096	.	.
SEMINAL VESICLE	ADENOCARCINOMA	1	0	0	0	0.7593	0.5048	0.5140	0.5185
SKELETAL MUSCLE	HAEMANGIOSARCOMA	0	2	0	0	0.7672	0 2573	.	.
SK N	BASAL CELL CARCINOMA	0	1	0	0	0.5163	0 5096	.	.
	BASAL CELL TUMOUR	0	1	0	0	0.5163	0 5096	.	.
	FIBROMA	1	2	3	1	0.5214	0 5146	0.3380	0.2716
	FIBROSARCOMA	0	2	0	2	0.2295	0 2573	.	0.2762
	HAEMANGIOMA	0	1	1	1	0.3177	0 5096	0.5189	0.5234
	HAEMANGIOSARCOMA	1	0	0	0	0.7628	0 5096	0.5189	0.5234
	KERATOACANTHOMA	5	3	4	1	0.9345	0 6422	0.5237	0.9093
	L POMA	1	0	1	0	0.6453	0 5096	0.2668	0.5234
	MALIGNANT SCHWANNOMA	0	1	1	0	0.5185	0 5096	0.5189	.
	SQUAMOUS CELL CARC N	0	0	0	1	0.2605	.	.	0.5234
	SQUAMOUS CELL PAPILL	0	2	1	0	0.7082	0 2573	0.5189	.
SPLEEN	HAEMANGIOMA	0	1	0	1	0.3298	0 5096	.	0.5234

File Name: NDA205832Carcin.doc

Organ Name	Tumor Name	0 mg Control N=60	2.5 mg Low N=60	5 mg Med N=60	10 mg High N=60	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
	HAEMANGIOSARCOMA	1	0	0	0	0.7593	0 5048	0.5140	0.5185
STOMACH	LEIOMYOSARCOMA	1	0	0	0	0.7628	0 5096	0.5189	0.5234
	MALIGNANT SCHWANNOMA	0	1	0	0	0.5163	0 5096	.	.
TA L	ADNEXAL POLYP	1	0	0	0	0.7593	0 5048	0.5140	0.5185
TESTES	INTERSTITIAL (LEYDIG	1	0	1	1	0.4300	0 5096	0.2668	0.2716
THYMUS	THYMOMA	1	0	4	0	0.5411	0 5048	0.1995	0.5185
THYROIDS	C-CELL ADENOMA	8	11	7	7	0.7899	0 3229	0.5472	0.5622
	C-CELL CARCINOMA	1	2	1	1	0.5888	0 5146	0.2668	0.2716
	FOLLICULAR CELL ADEN	4	2	3	0	0.9652	0 6700	0.5315	0.9495

Table 5B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Female Rats

Organ Name	Tumor Name	0 mg Cont N=60	2.5 mg Low N=60	5 mg Med N=60	10 mg High N=60	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
ADRENALS	CORTICAL ADENOMA	2	1	0	0	0.9436	0.5360	0.7622	0.7713
	MALIGNANT PHAEOCHROM	1	0	0	0	0.7630	0.5238	0.5098	0.5192
	PHAEOCHROMOCYTOMA	1	1	1	0	0.7446	0.2720	0.2574	0.5192
BRAIN	GRANULAR CELL TUMOUR	0	1	1	1	0.3081	0.5238	0.5098	0.5192
CAECUM	LEIOMYOSARCOMA	0	0	0	1	0.2559	.	.	0.5192
CLITORAL GLANDS	SQUAMOUS CELL CARCIN	1	0	1	0	0.6362	0.5238	0.2574	0.5192
DUODENUM	LEIOMYOMA	0	0	1	0	0.5024	.	0.5098	.
H-POIETIC TUMOU	HISTIOCYTIC SARCOMA	0	0	1	0	0.5047	.	0.5146	.
	MALIGNANT LYMPHOMA	0	0	1	2	0.0643	.	0.5146	0.2672
HEAD	CARCINOMA, ZYMBAL'S	0	0	0	1	0.2559	.	.	0.5192
	SQUAMOUS CELL CARCIN	1	0	0	1	0.5071	0.5238	0.5098	0.2672
HEART	PARAGANGLIOMA	0	0	0	1	0.2559	.	.	0.5192
LN MESENTERIC	HAEMANGIOMA	1	2	5	1	0.5110	0.5360	0.1164	0.2672
	HAEMANGIOSARCOMA	0	0	0	1	0.2559	.	.	0.5192
MAMMARY AREAS	MAMMARY ADENOCARCINO	8	2	1	0	0.9998	0.9620	0.9844	0.9975
	MAMMARY ADENOMA	1	2	4	0	0.6963	0.5360	0.2000	0.5192
	MAMMARY F BROADENOMA	19	17	15	9	0.9899	0.6575	0.7544	0.9821
OESOPHAGUS	HAEMANGIOMA	0	1	0	0	0.5024	0.5238	.	.
OVARIES	GRANULOSA CELL TUMOU	3	2	0	6	0.1025	0.5455	0.8858	0.2841
	LUTEOMA	0	1	0	0	0.5024	0.5238	.	.
	MALIGNANT GRANULOSA	2	0	1	0	0.8543	0.7756	0.5149	0.7713
PANCREAS	ISLET CELL ADENOMA	0	1	0	0	0.5024	0.5238	.	.
	ISLET CELL CARCINOMA	1	1	0	0	0.8206	0.2720	0.5098	0.5192
PARATHYROIDS	CH EF CELL ADENOMA	0	0	0	1	0.2559	.	.	0.5192
PITUITARY	ADENOMA, PARS DISTA	44	42	42	44	0.5445	0.8458	0.7652	0.6184
	ADENOMA, PARS NTERM	0	0	1	0	0.5024	.	0.5098	.
	CARCINOMA, PARS DIST	1	0	0	0	0.7630	0.5238	0.5098	0.5192
SALIVARY GLANDS	ADENOCARCINOMA	0	0	0	1	0.2559	.	.	0.5192
SKELETAL MUSCLE	F BROSARCOMA	0	1	0	0	0.5024	0.5238	.	.
	HAEMANGIOSARCOMA	0	1	0	0	0.5024	0.5238	.	.
SK N	F BROMA	0	0	1	0	0.5024	.	0.5098	.
	KERATOACANTHOMA	0	1	0	0	0.5024	0.5238	.	.
	LIPOMA	1	1	0	0	0 8206	0.2720	0.5098	0.5192
	SQUAMOUS CELL CARCIN	0	0	0	1	0.2559	.	.	0.5192
	ZYMBALS GLAND ADENOM	0	0	1	0	0.5047	.	0.5146	.
STOMACH	MALIGNANT SCHWANNOMA	0	0	0	1	0 2559	.	.	0.5192
	SQUAMOUS CELL PAPILL	0	1	0	0	0.5024	0.5238	.	.
THYMUS	THYMOMA	4	6	4	6	0.3702	0.4330	0.3910	0.4208
THYROIDS	C-CELL ADENOMA	10	11	12	3	0.9785	0.5756	0.4252	0.9692
	C-CELL CARCINOMA	0	1	1	0	0 5071	0.5238	0.5098	.
	FOLLICULAR CELL ADEN	1	2	0	0	0.8955	0.5360	0.5098	0.5192
	FOLLICULAR CELL CARC	0	1	0	0	0.5024	0.5238	.	.
UTERINE CERVIX	SQUAMOUS CELL CARCIN	1	0	0	0	0.7630	0.5238	0.5098	0.5192
UTERUS	ENDOMETRIAL ADENOCAR	1	2	0	0	0.8955	0.5360	0.5098	0.5192
	ENDOMETRIAL ADENOMA	1	3	0	0	0.9320	0.3458	0.5098	0.5192
	ENDOMETRIAL POLYP	5	5	8	4	0.5997	0.4330	0.3297	0.5492
	LEIOMYOMA	0	0	0	1	0.2559	.	.	0.5192
	MALIGNANT SCHWANNOMA	1	1	3	0	0 6503	0.2720	0.3312	0.5192

Organ Name	Tumor Name	0 mg Cont N=60	2.5 mg Low N=60	5 mg Med N=60	10 mg High N=60	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
	SQUAMOUS CELL CARCIN	1	0	1	0	0.6374	0.5238	0.2623	0.5192

Table 6A: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Male Mice

Organ Name	Tumor Name	0 mg Cont N=66	5 mg Low N=66	15 mg Med N=66	30 mg High N=66	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
ABDOMEN	OSTEOSARCOMA	1	0	0	0	0.7219	0.4951	0.4951	0.3882
ADRENALS	CORTICAL ADENOMA	1	2	1	1	0.4949	0.4854	0.7427	0.6316
	PHAEOCHROMOCYTOMA	1	1	0	1	0.4827	0.7476	0.4951	0.6286
	SUBCAPSULAR CELL ADE	3	3	2	0	0 9094	0 6517	0.4907	0.7762
BONE	OSTEOSARCOMA	0	0	0	1	0.1809	.	.	0.3953
COLON	ADENOCARCINOMA	1	1	0	0	0.7713	0.7476	0.4951	0.3882
DUODENUM	ADENOMA	0	1	0	0	0.4492	0.4951	.	.
H-POIETIC TUMOU	MALIGNANT LYMPHOMA	3	4	5	5	0 0990	0.4893	0.3685	0.1697
	MYELOID CELL LEUKAEM	1	0	0	0	0.7219	0.4951	0.4951	0.3882
HARDERIAN GLAND	ADENOCARCINOMA	0	1	0	0	0.4492	0.4951	.	.
	ADENOMA	5	7	10	6	0.1385	0 3938	0.1320	0.2379
KIDNEYS	TUBULAR ADENOMA	0	3	0	0	0.7720	0.1178	.	.
LIVER	HAEMANGIOMA	0	1	0	1	0 2239	0.4951	.	0.3882
	HAEMANGIOSARCOMA	0	3	2	0	0 5896	0.1178	0.2427	.
	HEPATOCELLULAR ADENO	13	9	12	6	0 6511	0.7479	0.4781	0.6852
	HEPATOCELLULAR CARCI	1	4	2	1	0 5279	0.1813	0.4926	0.6286
LUNGS + BRONCHI	BRONCHIOLOALVEOLAR A	13	14	22	8	0 3980	0.4770	0.0419	0.4663
		5	8	7	2	0.7357	0 2652	0.3803	0.5645
PANCREAS	ISLET CELL ADENOMA	0	1	0	0	0.4492	0.4951	.	.
PITUITARY	ADENOMA, PARS DISTA	0	1	0	0	0.4492	0.4951	.	.
SKELETAL MUSCLE	F BROSARCOMA	0	0	0	1	0.1809	.	.	0.3953
	SARCOMA, NOS	1	1	0	0	0.7707	0.7524	0.4951	0.3882
SK N	F BROSARCOMA	0	0	1	0	0.4521	.	0.5000	.
SPLEEN	HAEMANGIOSARCOMA	0	1	0	0	0.4468	0 5000	.	.
TESTES	INTERSTITIAL (LEYDIG	4	3	0	2	0.7234	0.4783	0.9363	0.4360
THYMUS	MALIGNANT THYMOMA	1	0	0	0	0.7219	0.4951	0.4951	0.3882

Table 6B: Tumor Rates and P-Values for Dose Response Relationship and Pairwise Comparisons – Female Mice

Organ Name	Tumor Name	0 mg Cont N=66	5 mg Low N=66	15 mg Med N=66	30 mg High N=66	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
ABDOMEN	OSTEOSARCOMA	0	0	1	0	0.4919	.	0.4948	.
ADRENALS	PHAEOCHROMOCYTOMA	1	0	0	0	0.7297	0.4737	0.4845	0.4624
	SUBCAPSULAR CELL ADE	1	0	1	0	0 6036	0.4787	0.7421	0.4674
CAECUM	ADENOMA	0	1	0	0	0.4891	0.4787	.	.
DUODENUM	ADENOCARCINOMA	0	0	1	0	0.4891	.	0.4896	.
H-POIETIC TUMOU	HISTIOCYTIC SARCOMA	3	4	3	3	0.4858	0.4530	0.6415	0.5984
	MALIGNANT LYMPHOMA	**14**	**17**	**17**	**11**	**0.7138**	**0.3107**	**0.3343**	**0.5669**
	MYELO D CELL LEUKAEM	1	1	0	0	0.7958	0.7256	0.4845	0.4624
	PLASMA CELL LYMPHOMA	**0**	**0**	**0**	**1**	**0.2337**	**.**	**.**	**0.4674**
	COMB LYMPHOMA	**14**	**17**	**17**	**12**	**0.6245**	**0.3107**	**0.3343**	**0.4708**
HARDERIAN GLAND	ADENOMA	8	4	2	7	0.4845	0.7667	0.9474	0.4111
LIVER	HAEMANGIOSARCOMA	1	0	1	2	0.1676	0.4787	0.7421	0.4593
	HEPATOCELLULAR ADENO	1	1	1	2	0 2558	0.7256	0.7369	0.4517
	HEPATOCELLULAR CARCI	0	2	0	0	0.7404	0 2265	.	.
LUNGS + BRONCHI	BRONCHIOLOALVEOLAR A	14	8	9	8	0 8210	0.8258	0.7997	0.8258
		6	1	3	4	0.5152	0 9223	0.7252	0.5473
MAMMARY AREAS	ADENOMYOEPITHELIOMA	0	0	0	1	0 2337	.	.	0.4674
	MAMMARY ADENOACANTHO	**0**	**1**	**0**	**1**	**0.2886**	**0.4787**	**.**	**0.4674**
	MAMMARY ADENOCARCINO	**0**	**2**	**1**	**4**	**0.0270***	**0.2318**	**0.4948**	**0.0465***
	MAMMARY ADENOMA	**1**	**0**	**0**	**0**	**0.7337**	**0.4787**	**0.4896**	**0.4674**
	COMB MAMMARY	**1**	**3**	**1**	**5**	**0.0581**	**0 2849**	**0.7474**	**0.0791**
OVAR ES	CYSTADENOMA	2	0	0	2	0 3466	0.7310	0.7421	0.6406
	HAEMANGIOMA	0	0	0	1	0 2378	.	.	0.4731
	LUTEOMA	2	2	4	2	0.4273	0 6586	0.3289	0.6498
	SERTOL FORM TUBULAR	2	1	2	0	0 8076	0.4678	0.6756	0.7191
	THECAL CELL TUMOUR	1	1	0	0	0.7992	0.7310	0.4896	0.4674
PANCREAS	ISLET CELL ADENOMA	0	1	0	0	0.4891	0.4787	.	.
	ISLET CELL CARC NOMA	0	0	1	0	0.4919	.	0.4948	.
PITUITARY	ADENOMA, PARS DISTA	4	1	0	1	0 9032	0.7916	0.9362	0.7839
SKELETAL MUSCLE	OSTEOSARCOMA	0	2	0	0	0.7377	0 2318	.	.
SK N	FIBROMA	1	0	0	0	0.7297	0.4737	0.4845	0.4624
	FIBROSARCOMA	2	0	1	0	0 8165	0.7310	0.4922	0.7191
	FIBROUS HISTIOCYTOMA	1	0	0	0	0.7337	0.4787	0.4896	0.4674
	RHABDOMYOSARCOMA	1	0	0	0	0.7297	0.4737	0.4845	0.4624
	SQUAMOUS CELL CARCIN	0	1	0	0	0.4891	0.4787	.	.
SPLEEN	FIBROUS HISTIOCYTOMA	0	1	0	0	0.4891	0.4787	.	.
STOMACH	SQUAMOUS CELL PAPILL	0	1	0	0	0.4891	0.4787	.	.
THORAX	RHABDOMYOSARCOMA	0	1	0	0	0.4891	0.4787	.	.
THYMUS	HAEMANGIOSARCOMA	1	0	0	0	0.7297	0.4737	0.4845	0.4624
THYROIDS	FOLLICULAR CELL ADEN	1	0	0	0	0.7297	0.4737	0.4845	0.4624
UR NARY BLADDER	MESENCHYMAL TUMOUR	0	0	1	0	0.4891	.	0.4896	.
UTERINE CERVIX	POLYP	0	1	2	0	0.4266	0.4787	0.2423	.
UTERUS	ENDOMETRIAL ADENOCAR	0	1	1	0	0.4781	0.4787	0.4948	.
	ENDOMETRIAL ADENOMA	1	1	0	0	0.7958	0.7256	0.4845	0.4624
	ENDOMETRIAL POLYP	12	7	6	3	0 9862	0.7798	0.8775	0.9763
	ENDOMETRIAL STROMAL	0	2	0	0	0.7404	0 2265	.	.

File Name: NDA205832Carcin.doc

Organ Name	Tumor Name	0 mg Cont N=66	5 mg Low N=66	15 mg Med N=66	30 mg High N=66	P-Value			
						Dos Resp	C vs. L	C vs. M	C vs. H
	GRANULAR CELL TUMOUR	1	0	1	1	0 3856	0.4787	0.7474	0.7191
	HISTIOCYTIC SARCOMA	2	2	0	2	0.4847	0.6506	0.7369	0.6417
	LEIOMYOMA	5	2	2	5	0 3251	0.7364	0.7558	0.5456
	LEIOMYOSARCOMA	0	0	1	0	0.4919	.	0.4948	.
	MALIGNANT SCHWANNOMA	0	1	0	0	0.4891	0.4787	.	.
VAG NA	POLYP	0	2	1	0	0.6688	0 2265	0.4948	.

Figure 1A: Kaplan-Meier Survival Functions for Male Rats

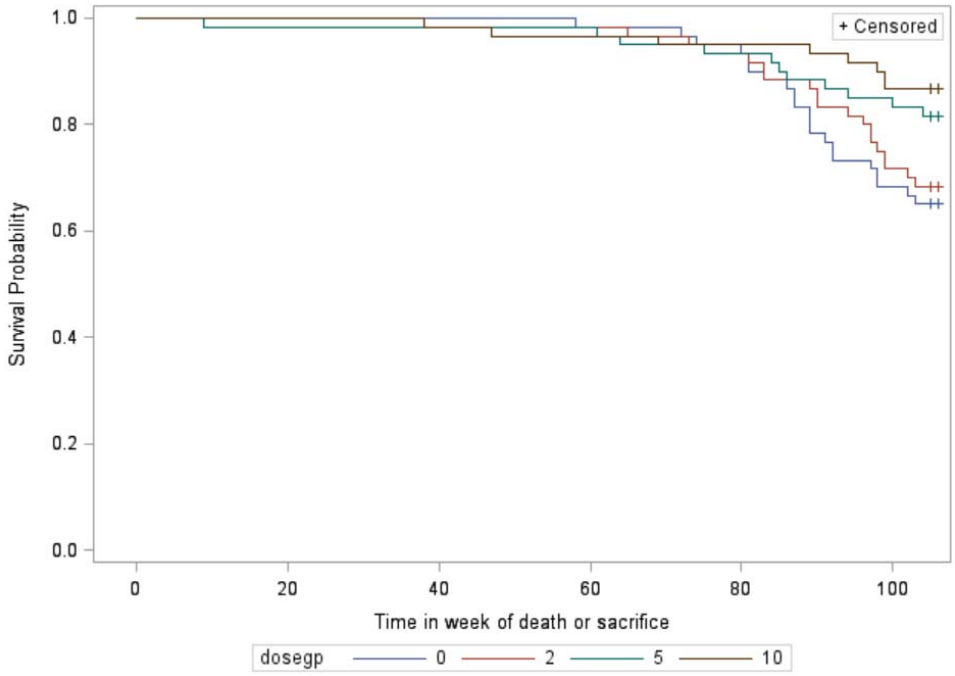


Figure 1B: Kaplan-Meier Survival Functions for Female Rats

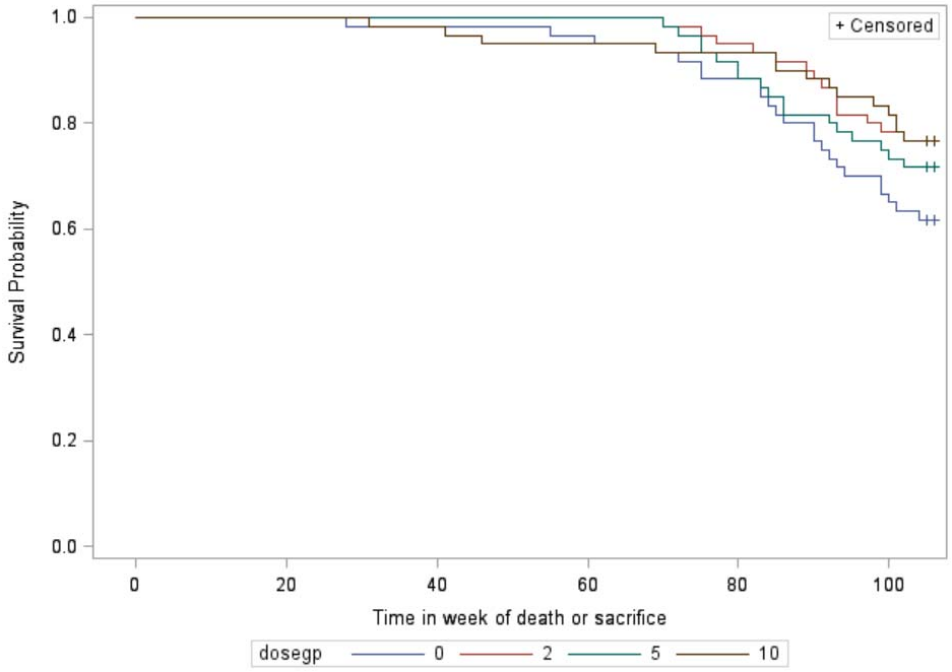


Figure 2A: Kaplan-Meier Survival Functions for Male Mice

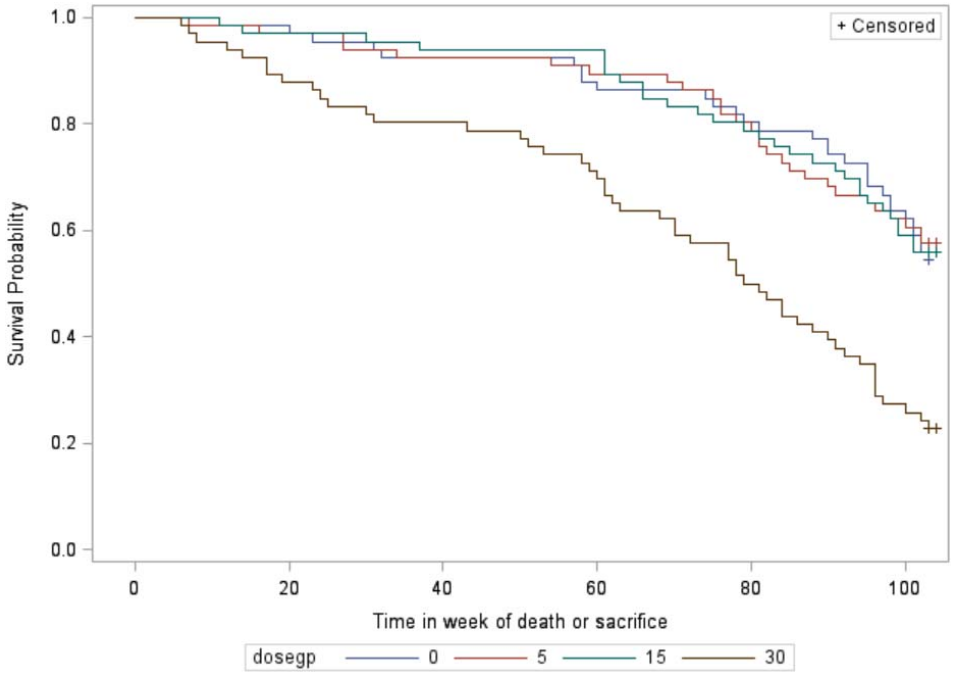
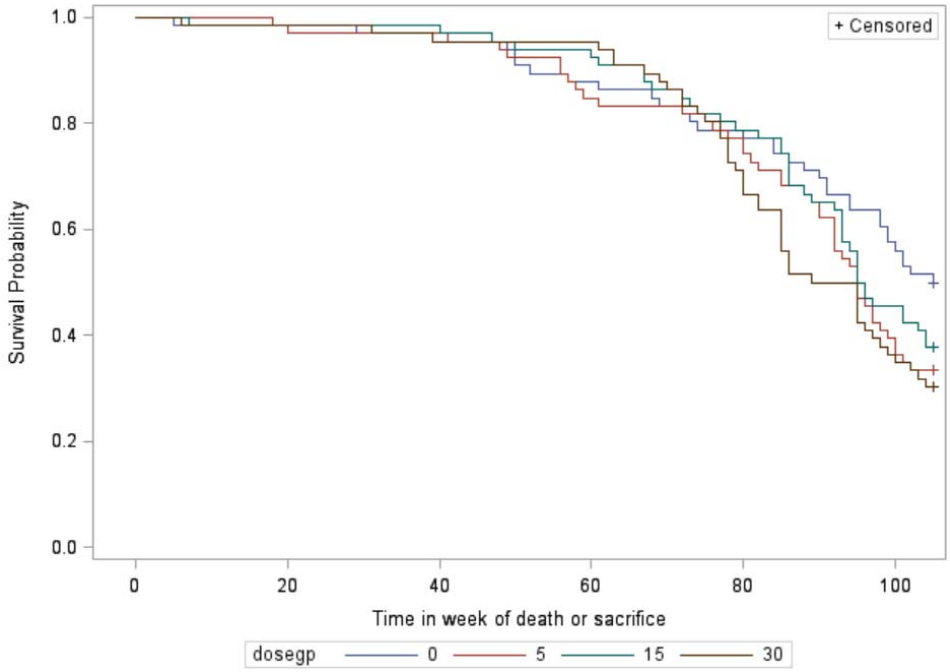


Figure 2B: Kaplan-Meier Survival Functions for Female Mice



8 References

1. Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J.Wahrendorf, "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva,* 311-426, 1980.
2. Bailer AJ, Portier CJ (1988). "Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples." *Biometrics*, 44, 417-431.
3. Bieler, G. S. and Williams, R. L. (1993). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity". *Biometrics* 49, 793-801.
4. Tarone RE, "Test for trend in life table analysis", *Biometrika* 1975, 62: 679-82
5. Lin K.K. and Rahman M.A.," Overall false positive rates in tests for linear trend in tumor incidence in animal carcinogenicity studies of new drugs", *Journal of Biopharmaceutical Statistics,* 8(1), 1-15, 1998.
6. Haseman, J, "A re-examination of false-positive rates for carcinogenesis studies", *Fundamental and Applied Toxicology*, 3: 334-339, 1983.
7. Guidance for Industry. Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Statues of Pharmaceuticals (Draft Guidance). U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), May 2001.

--EOF--

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

FENG ZHOU
07/17/2014

KARL K LIN
07/17/2014
Concur with review

STATISTICS FILING CHECKLIST FOR NDA 205-832

NDA Number: 205-832 **Applicant:** Boehringer Ingelheim **Stamp Date:** May 2, 2014

Drug Name: Nintedanib **NDA Type:** Priority

On **initial** overview of the NDA application for RTF: **Studies 1199.32, 1199.34, 1199.30**

	Content Parameter	Yes	No	NA	Comments
1	Index is sufficient to locate necessary reports, tables, data, etc.	X			
2	ISS, ISE, and complete study reports are available (including original protocols, subsequent amendments, etc.)	X			
3	Safety and efficacy were investigated for gender, racial, and geriatric subgroups investigated (if applicable).	X			
4	Data sets in EDR are accessible and do they conform to applicable guidances (e.g., existence of define.pdf file for data sets).	X			

IS THE STATISTICAL SECTION OF THE APPLICATION FILEABLE? Yes

Comment:

Please identify and list any potential review issues to be forwarded to the Applicant for the 74-day letter.

Content Parameter (possible review concerns for 74-day letter)	Yes	No	NA	Comment
Designs utilized are appropriate for the indications requested.	X			
Endpoints and methods of analysis are specified in the protocols/statistical analysis plans.	X			
Interim analyses (if present) were pre-specified in the protocol and appropriate adjustments in significance level made. DSMB meeting minutes and data are available.			X	
Appropriate references for novel statistical methodology (if present) are included.			X	
Safety data organized to permit analyses across clinical trials in the NDA/BLA.			X	
Investigation of effect of dropouts on statistical analyses as described by applicant appears adequate.	X			

STATISTICS FILING CHECKLIST FOR NDA 205-832

Brief Summary of Pivotal Studies

Trial ID	Design	Treatment/ Sample Size	Endpoint	Preliminary Findings
1199.32	A 52-weeks, placebo-controlled, multiregional parallel study	Nintedanib 150 mg/309 Placebo/206	Primary: Annual rate of decline in FVC (mL/year). Key Secondary: Time to first acute IPF exacerbation, Change from baseline in SGRQ total score at Week 52	Primary: Annual rate of decline in FVC (mL/year): Nintedanib vs Placebo (Difference = 125; 95% CI 78, 173). Key Secondary: Time to first acute IPF exacerbation: Nintedanib vs Placebo (Hazard Ratio = 1.15; 95% CI 0.54, 2.42) Change from baseline in SGRQ total score at Week 52: Nintedanib vs Placebo (Difference = -0.05; 95% CI -2.50, 2.40)
1199.34	A 52-weeks, placebo-controlled, multiregional parallel study	Nintedanib 150 mg/331 Placebo/220	Primary: Annual rate of decline in FVC (mL/year) Key Secondary: Time to first acute IPF exacerbation, Change from baseline in SGRQ total score at Week 52	Primary EP: Annual rate of decline in FVC (mL/year): Nintedanib vs Placebo (Difference = 94; 95% CI 45, 143) Key Secondary EP:Time to first acute IPF exacerbation: Nintedanib vs Placebo(Hazard Ratio = 0.38; 95% CI 0.19, 0.77) Change from baseline in SGRQ total score at Week 52: Nintedanib vs Placebo (Difference = -2.69; 95% CI -4.95, -0.40)

Data source location: \\CDSESUB5\EVSPROD\NDA205832\205832.enx

Note that the results from the Phase II Study 1199.30 will also be described in the review.

STATISTICS FILING CHECKLIST FOR NDA 205-832

Additional information regarding the data:

	Information regarding the data	Comments
1	Dataset location	\\Cdsesub1\evsprod\NDA205832\0000\m5\datasets\1199-0030 \\Cdsesub1\evsprod\NDA205832\0000\m5\datasets\1199-0032 \\Cdsesub1\evsprod\NDA205832\0000\m5\datasets\1199-0034
2	Dataset structure (e.g., SDTM or ADaM)	Legacy analysis datasets
3	Based on the analysis datasets, can results of the primary endpoint(s) be reproduced? (Yes or No)	Yes
4	List the dataset(s) that contains the primary endpoint(s)	indpft.xpt , indsgrq.xpt, and indtte
5	Are there any concerns about site(s) that could lead to inspection? If so, list of site(s) that needs inspection and rationale	None identified (results of the primary endpoints were consistent across different regions) In Study 1199.32, largest # of subjects/site is 20 (site 86001 Dr. Zuojun Xu in China) and only five sites have more than 15 subjects. Total # of sites is 98. In Study 1199.34, largest # of subjects/site is 30 (site 86052 Dr. Huiping Li in China) and only six sites have more than 15 subjects. Total # of sites is 107.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

YONGMAN KIM
06/06/2014

DAVID M PETULLO
06/07/2014