CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

22-527

STATISTICAL REVIEW(S)

ADDENDUM

Date: August 10, 2010
From: Sharon Yan (HFD-710)
To: File NDA 22,527 Fingolimod
Subject: Statistical analysis of new or newly enlarged T2 lesions – Study D2302

This addendum is intended to replace the **New or newly enlarged T2 lesions at Month 12** in **Section 3.1.2.5.2 Efficacy Results of Key Secondary Endpoints** of the **Statistical Review and Evaluation.**

Background

New or newly enlarged T2 lesion mean is a key secondary efficacy endpoint for Study D2302. In the original submission, the sponsor reported that statistical significance in treatment difference in the mean of new or newly enlarged T2 lesion count was only reached for the FTY720 1.25 mg group, as compared to Interferon beta-1a group. In an addendum filed on 22 November 2009, the sponsor claimed that new or newly enlarged T2 lesions were not read correctly, and statistical significance in treatment difference was reached for both FTY720 dose groups after correcting the mistakes. In addition, results from 18 of the subjects could not be included in the analysis filed in the addendum because correct lesion counts for the subjects were not available. Those 18 subjects had MRI scans performed when they discontinued the study and at Month 12. Originally, the Month 12 scan was compared to the one when they discontinued the study, and new T2 lesion count at Month 12 as compared to baseline scan was not recorded.

After FDA's review of the addendum and a discussion with the sponsor, it was agreed that the correct count of the lesions should be directly read from the original scans, not derived from the readings, and the sponsor would have the original scans, including the scans of those 18 subjects, re-read to obtain the correct lesion count. The sponsor submitted the results and data from the recount to FDA in July 2010.

Due to the reasons stated above, analysis of new or newly enlarged T2 lesions by the reviewer was delayed and not included in the original statistical review. Statistical analysis of new or newly enlarged T2 lesions was performed by the reviewer after the recount data were submitted. The following sections present the statistical analysis results obtained by the sponsor and results obtained by the reviewer using the data from the recount.

New Results from Sponsor's Analysis

The number of new or newly enlarged T2 lesions at Month 12 was compared between treatment groups using a negative binomial regression model adjusting for the same covariates used in the primary efficacy analysis (treatment, country, baseline number of relapses in the previous 2 years, and baseline EDSS), and the results are shown in Table 1.

Table 1 Number of new or newly enlarged T2 lesions at Month 12 – Study D2302 (Source: Table 2-1 of Response to FDA Request for Information, 28 June 2010)

	FTY720 1.25mg N=420	FTY720 0.5mg N=429	Interferon beta-1a i.m. N=431
n	356	380	365
Mean (SD)	1.6 (3.23)	1.6 (3.16)	2.6 (5.50)
Median	1.0	0.0	1.0
Range	0 - 42	0 - 23	0 - 56
P-value for treatment comparison of FTY720 vs. Interferon beta-1a i.m. (negative binomial regression with covariates)	0.002*	0.002*	–

n=the number of patients with evaluable MRI at baseline and Month 12
P-value is calculated using a negative binomial regression model adjusting for treatment, country, baseline number of relapses in the previous 2 years, and baseline EDSS.
* Indicates two-sided statistical significance at 0.05 level.
Source: [Table 14.2-1.31c]

For the ITT population both the FTY720 1.25 mg and FTY720 0.5 mg treatment groups had a lower mean number of new or newly enlarged T2 lesions at Month 12 compared to the interferon beta-1a i.m. group, reaching statistical significance for both doses (p=0.002). The sponsor reported that the results were supported by the analysis in the per-protocol population, where the mean number of new or newly enlarged T2 lesions at Month 12 compared to the interferon beta-1a i.m. group were lower for both FTY720 doses (p=0.009 for FTY720 1.25 mg and p=0.002 for FTY720 0.5 mg). The 18 patients previously excluded because their 12-month T2 data were not compared to baseline are now included in the analysis after the recount.

The requested sensitivity analysis using the recount of new/newly enlarged T2 lesions with the MRI at the time of study drug end-point carried forward are provided in the following table.

Table 2 Number of new or newly enlarged T2 lesions at Month 12 (Study drug endpoint carry-forward, ITT population) (Source: Table 2-2 of Response to FDA Request for Information, 28 June 2010)

	FTY720 1.25mg N=420	FTY720 0.5mg N=429	Interferon beta-1a i.m. N=431
n	379	393	385
Mean (SD)	1.5 (3.20)	1.5 (3.09)	2.6 (5.48)
Median	1.0	0.0	1.0
Range	0 - 42	0 - 23	0 - 56
P-value for treatment comparison of FTY720 vs. Interferon beta-1a i.m. (negative binomial regression with covariates)	0.001*	<0.001*	–

n=the number of patients with evaluable MRI at baseline and Month 12 for patients who completed the study or at baseline and study drug end point for patients who did no complete the study.
P-value is calculated using a negative binomial regression model adjusting for treatment, country, baseline number of relapses in the previous 2 years, and baseline EDSS.
* Indicates two-sided statistical significance at 0.05 level.
Source: [pt-table-14-2-1-31d-response-20100504-fda]

The proportion of patients free of new or newly enlarged T2 lesions at Month 12 was higher in the FTY720 0.5 mg group (54.5%) compared to the FTY720 1.25 mg (48.3%) and interferon beta-1a i.m. treatment groups (46.0%). Only the comparison of FTY720 0.5 mg group vs. the interferon beta-1a i.m. group reached statistical significance (p=0.012).

Results from Reviewer's Analysis

The submitted data included 1171 subjects with MRI T2 new or newly enlarged lesion count. One subject in the interferon β-1a group who had missing baseline value is excluded from the primary analysis for T2 lesions. In the sponsor's analysis, 1101 subjects were included in the Month 12 analysis and 1157 subjects were included in the study drug endpoint carry-forward analysis. It is not clear why some subjects were not included in the sponsor's analysis.

Many subjects had multiple MRI values, mostly because they had MRI performed at discontinuation, at the follow-up visit 3 months after the drug discontinuation, and at the Month 12 visit. For those subjects, the MRI value at or near the drug discontinuation was used. Therefore, the results from the reviewer's analysis should be compared to sponsor's results shown in Table 2.

Table 3 New or newly enlarged T2 lesions at Month 12 – Study drug endpoint carry forward (Source: reviewer's analysis)

	FTY720 1.25 mg N=385	FTY720 0.5 mg N=399	IFN β-1a N=386
Mean (SD) new or enlarged T2			
Unadjusted (observed)	1.58 (3.26)	1.63 (3.30)	2.61 (5.48)
Adjusted	1.65	1.62	2.62
95% CI	(1.35, 2.01)	(1.33, 1.97)	(2.08, 3.07)
p-value	.0017	.0007	

The mean number of new or newly enlarged T2 lesions found by the reviewer is slightly larger than the one obtained by the sponsor for all treatment groups.

Cc: Toure (Project Manager) (HFD-120)
Fitter (HFD-120)
Bastings (HFD-120)
Katz (HFD-120)
Jin (HFD-710)
Mahjoob (HFD-710)
Hung (HFD-710)
Patrician (HFD-700)

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22527	ORIG-1	NOVARTIS PHARMACEUTICALS CORP	FINGOLIMOD HCL ORAL CAPSULES

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

/s/

XIAORONG YAN
08/10/2010

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	22527 / Serial Number 000-
Drug Name:	FTY720 (Fingolimod Hydrocloride)
Indication(s):	Multiple Sclerosis
Applicant:	Novartis
Date(s):	Rolling Submission Completion Date: 12/21/2009
	PDUFA Due Date: 6/21/2010
Review Priority:	Priority Review
Biometrics Division:	Division of Neurology
Statistical Reviewer:	Sharon Yan
Concurring Reviewers:	Kun Jin
	Kooros Mahjoob
Medical Division:	Neurology
Clinical Team:	Heather Fitter, M.D., Clinical Reviewer
	Lourdes Villalba, M.D., Safety Reviewer
	Eric Bastings, M.D., Deputy Director
	Russell Katz, M.D., Director
Project Manager:	Hamet Toure, Pharm.D., Lt., Project Manager

Table of Contents

1. EXECUTIVE SUMMARY

This NDA submission is to obtain marketing authorization for FTY720 (fingolimod hydrochloride) in the treatment of relapsing multiple sclerosis (MS). The proposed recommended dose is 0.5 mg once-daily administered orally. The proposed indication is as disease-modifying therapy for treatment of patients with relapsing MS to reduce the frequency of relapses and to delay the accumulation of physical disability.

1.1 Conclusions and Recommendations

The two Phase III studies in patients with relapsing MS presented evidence that FTY720, at once daily oral doses of 1.25 mg and 0.5 mg, is efficacious for the treatment of relapsing MS based on reduction in annual relapse rate. Patients in both of the FTY720 dose groups in the 2-year study D2301 were also found having significantly slower progression of disability, but such result was not replicated in the 1-year study D2302.

1.2 Brief Overview of Clinical Studies

The Phase III clinical development program of FTY720 in relapsing MS included 3 Phase III studies (D2301, D2302, D2309), all evaluating the efficacy and safety of FTY720 at once daily oral doses of 0.5 mg and 1.25 mg. Two of the studies (D2301, D2302), which constitute the pivotal program, are completed and included in the current submission. The third one is still ongoing at time of submission of this NDA.

1.3 Statistical Issues and Findings

No major statistical issues were found.

2. INTRODUCTION

Two large adequate and well-controlled Phase III studies contributed the bulk of the efficacy data in this submission; one 2-year, placebo-controlled study, and a one-year, active-controlled study employing interferon β-1a (IFN β-1a, Avonex®) as the active comparator. Another 6-month, Phase II placebo-controlled study is also submitted but not included in this review.

2.1 Overview

The Phase III clinical development program of FTY720 in relapsing MS included 3 Phase III studies (D2301, D2302, D2309), all evaluating the efficacy and safety of FTY720 at once daily oral doses of 0.5 mg and 1.25 mg. Two of these studies (D2301, D2302), which constituted the pivotal program, are completed and are included in the current submission:

- D2301: a 2-year, double-blind, placebo-controlled study in 1272 patients with relapsing remitting MS (RRMS) conducted globally (outside of the USA). The primary endpoint was annualized relapse rate. The aggregate ARR was significantly lower in both FTY720 groups compared with the placebo group. The magnitude of treatment effect (relapse reduction relative to placebo) was 60% for the 1.25 mg group and 54% in the 0.5 mg group with no significant difference between the two FTY720 doses.

- D2302: a 1-year, double-blind, double-dummy, active-controlled (once weekly 30µg intramuscular IFN β-1a, Avonex®) study in 1292 patients with RRMS conducted globally including US. The primary endpoint was annualized relapse rate. The ARR was significantly lower in both FTY720 groups compared with the IFN β-1a group, resulting in a relative reduction in the ARR of 38% for 1.25 mg group and 52% for 0.5 mg group, with no significant difference between the two FTY720 doses.

The third, still ongoing study (D2309), is a 2-year, double-blind, placebo-controlled study in approximately 1080 patients with RRMS, conducted mainly in the USA (recruitment completed).

All three studies included a long-term extension phase, which are still on-going. A phase II, 6-month placebo controlled MRI study is also included in the submission.

2.2 Data Sources

All documents reviewed for this NDA submission are in electronic form. The path to CDER Electronic Document Room for documents of this NDA is listed below:

\\Cdsesub1\evsprod\NDA022527

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study D2301

3.1.1.1 Description of the Study

The primary objective was to compare two doses of FTY720 (1.25 mg and 0.5 mg) with placebo and to demonstrate that at least 1.25 mg FTY720 was superior to placebo in terms of annualized relapse rate (ARR) in patients with RRMS treated for up to 24 months. The key secondary objective was to evaluate the effect of FTY720 1.25 mg and 0.5 mg relative to placebo on time to 3-month confirmed disability progression as measured by EDSS in patients treated for up to 24 months.

This was a 24-month, double-blind, randomized, multicenter, placebo-controlled, parallel-group study. Patients were randomized to receive FTY720 0.5 mg/day, FTY720 1.25 mg/day, or placebo for up to 24 months.

The study was conducted in 138 centers in 22 non-US countries. A total of 1250 patients were planned and 1272 patients were actually randomized. Patients enrolled in this study were diagnosed of MS based on 2005 revised McDonald criteria, were treatment naïve or previously treated, had a relapsing-remitting course with at least one documented relapse during the previous year or two documented relapses during the previous 2 years, prior to randomization, and had Expanded Disability Status Scale (EDSS) score of 0 to 5.5 inclusive.

3.1.1.2 Efficacy Variables

3.1.1.2.1 Primary Efficacy Endpoint

The primary endpoint of this study was the aggregate annualized relapse rate (ARR) at 24 months, which was defined as the number of relapses per year. Only confirmed relapses were considered for the primary analyses.

Relapse was defined as an appearance of a new neurological abnormality or worsening of previously stable or improving pre-existing neurological abnormality, separated by at least 30 days from the onset of a preceding clinical demyelinating event. The abnormality must have been present for at least 24 hours and have occurred in the absence of fever (< 37.5°C) or infection.

A relapse was confirmed when it was accompanied by an increase of at least half a step (0.5) on the EDSS, or an increase of 1 point on two different Functional Systems (FS) of the EDSS, or 2 points on one of the FS (excluding Bowel/Bladder or Cerebral FS), and it was confirmed by the Independent Evaluating Physician (examining neurologist).

The average ARR was calculated in two ways:

1. Group level (aggregate ARR)

The ARR of the treatment group was calculated by taking the total number of confirmed relapses for all patients in the treatment group divided by the total number of days on study for all patients in the group and multiplied by 365.25 to obtain the annual rate.

2. Patient level

The ARR for each patient was calculated as the total number of confirmed relapses divided by total number of days on study, multiplied by 365.25. The ARR for each treatment group was the mean of ARRs from all patients in the group.

3.1.1.2.2 Secondary Efficacy Endpoint

The key secondary endpoint was the time to 3-month confirmed disability progression assessed at 24 months.

Confirmed disability progression was defined as 1 point EDSS increase from baseline or 0.5 point increase if baseline EDSS was ≥ 5.5, confirmed 3 months later. A 3-month confirmed progression was defined as a 3-month sustained increase from baseline EDSS score. It means that every EDSS score obtained (scheduled or unscheduled) within a 3-month duration after the first progression should also meet the progression criteria.

Disability progression could only be confirmed at a scheduled visit in the absence of a relapse.

If a patient died due to MS after the start of a tentative disability progression event, then it would be considered as a confirmed progression. If a patient died due to MS before having progression, then the time to disability progression was to be censored using the date of death.

3.1.1.3 Statistical Analysis Methods

Efficacy analyses were to be performed on the ITT population. ITT patient population consisted of all patients who were randomized and received at least one dose of study medication. Patients were grouped according to the assigned treatment.

3.1.1.3.1 Analysis of the Primary Efficacy Variable

The primary null hypotheses to be tested were: 1) there was no difference in the aggregate annualized relapse rate (ARRs) between patients treated with the FTY720 1.25 mg versus placebo, and 2) there was no difference in the aggregate ARRs between patients treated with the FTY720 0.5 mg versus placebo.

The test of the null hypotheses was to be based on a negative binomial regression model using treatment group, country, number of relapses in previous 2 years, and baseline EDSS as covariates. Number of relapses in the previous 2 years and baseline EDSS were to be treated as continuous covariates in the model. Individual countries with small number of patients were to be pooled for analysis.

For the negative binomial regression, the response variable was the number of relapses for each patient and quadratic variance estimate was to be used. Log of time on study in years was to be used as the offset variable to account for the varying lengths of patients' time in the study. The ARR and its 95% confidence interval for each treatment group were to be estimated from the model.

For patients who prematurely discontinue study drug, the relapses collected after the study drug discontinuation were to be included in the analysis.

3.1.1.3.2 Analysis of the Key Secondary Efficacy Variable

The treatment groups were to be compared for time to disability progression using a log-rank test.

A patient was to be censored if the patient prematurely withdrew from the study or completed the study before the onset of a disability progression or before the progression could be confirmed if an onset had occurred. Therefore, any disability progression onset occurred after the 21 month visit was to be treated as censored in the analyses.

3.1.1.3.3 Multiplicity Adjustment

To control the overall type-I error rate of the study, a multiplicity adjustment was to be applied to the primary and key secondary endpoints.

There was one primary endpoint and one key secondary endpoint with two doses, which yielded a total of four comparisons. The testing was to be done in a hierarchical order as follows:

1. FTY720 1.25 mg vs. placebo testing treatment difference for aggregate ARR;
2. FTY720 0.5 mg vs. placebo testing treatment difference for aggregate ARR;

3. FTY720 1.25 mg vs. placebo testing treatment difference for time to 3-month confirmed disability progression;
4. FTY720 0.5 mg vs. placebo testing treatment difference for time to 3-month confirmed disability progression.

Each testing was to be performed at a significant level of 0.05 for these four ranked comparisons. However, the lower-ranked testing was to be performed only when every higher-ranked testing preceding it was statistically significant at 0.05.

3.1.1.4 Patient Results

3.1.1.4.1 Patient Disposition

A total of 1564 patients were screened for participation in this study. Of the 1272 patients who were randomized, 1033 (81.2%) completed the study, with the highest percentage of patients completing in the FTY720 0.5 mg group (86.8%) compared with 77.4% and 79.4% for the FTY720 1.25 mg and placebo groups, respectively. A similar pattern was seen for those who completed the study while on study drug: 81.2% in the FTY720 0.5 mg group compared with 69.2% and 72.5% in the FTY720 1.25 mg and placebo groups, respectively. Patient disposition for the randomized population is presented in Table 1.

Table 1 Patient Disposition – Study D2301 (Source: Table 10-2 of sponsor's Study Report)

	FTY720 1.25mg N=429 n (%)	FTY720 0.5mg N=425 n (%)	Placebo N=418 n (%)	Total N=1272 n (%)
Completed study	332 (77.4)	369 (86.8)	332 (79.4)	1033 (81.2)
On study drug [1]	297 (69.2)	345 (81.2)	303 (72.5)	945 (74.3)
Off study drug [2]	35 (8.2)	24 (5.6)	29 (6.9)	88 (6.9)
Discontinued from the study	97 (22.6)	56 13.2)	86 (20.6)	239 (18.8)
Subject withdrew consent	31 (7.2)	17 (4.0)	28 (6.7)	76 (6.0)
Adverse event(s)	22 (5.1)	13 (3.1)	18 (4.3)	53 (4.2)
Unsatisfactory therapeutic effect	13 (3.0)	6 (1.4)	25 (6.0)	44 (3.5)
Abnormal laboratory value(s)	20 (4.7)	9 (2.1)	1 (0.2)	30 (2.4)
Lost to follow-up	3 (0.7)	5 (1.2)	7 (1.7)	15 (1.2)
Protocol violation	5 (1.2)	5 (1.2)	4 (1.0)	14 (1.1)
Abnormal test procedure result(s)	2 (0.5)	1 (0.2)	1 (0.2)	4 (0.3)
Death	1 (0.2)	0 (0.0)	2 (0.5)	3 (0.2)
Discontinued study drug	131 (30.5)	80 (18.8)	115 (27.5)	326 (25.6)
Subject withdrew consent	30 (7.0)	17 (4.0)	31 (7.4)	78 (6.1)
Adverse event(s)	31 (7.2)	15 (3.5)	24 (5.7)	70 (5.5)
Unsatisfactory therapeutic effect	18 (4.2)	8 (1.9)	36 (8.6)	62 (4.9)

Abnormal laboratory value(s)	32 (7.5)	20 (4.7)	5 (1.2)	57 (4.5)
Protocol violation	8 (1.9)	8 (1.9)	5 (1.2)	21 (1.7)
Lost to follow-up	2 (0.5)	6 (1.4)	5 (1.2)	13 (1.0)
Abnormal test procedure result(s)	6 (1.4)	3 (0.7)	3 (0.7)	12 (0.9)
Administrative problems	3 (0.7)	3 (0.7)	4 (1.0)	10 (0.8)
Death	1 (0.2)	0 (0.0)	2 (0.5)	3 (0.2)

[1] 'On study drug': Patients who took study drug until the study completion.
[2] 'Off study drug': Patients who completed the study but discontinued study drug prematurely.

Discontinuations from study drug were mostly common for safety reasons, i.e., adverse events and abnormal laboratory values, when taken together. The percentage of patients discontinuing for adverse events was lower in the FTY720 0.5 mg treatment group compared with the FTY720 1.25 mg and placebo treatment groups. The percentage of patients discontinuing for abnormal laboratory values was higher in the FTY720 treatment groups compared with the placebo group; of the two FTY720 groups, the percentage was higher in 1.25 mg group vs. the 0.5 mg group. Patients in the FTY720 0.5 mg treatment group discontinued from study drug due to withdrawal of consent less often compared with the FTY720 1.25 treatment group and placebo. Patients in the placebo group discontinued study drug due to unsatisfactory therapeutic effect at least twice as often compared with patients in the two FTY720 treatment groups.

Of 326 patients who discontinued study drug, 88 patients remained in the study and completed the abbreviated schedule of assessments through the Month 24 visit.

3.1.1.4.2 Baseline demographic characteristics

The study population was consistent with a population of RRMS patients in that approximately two thirds were female (69.9% female vs. 30.1% male), the majority (95.4%) were Caucasian, and the mean (SD) age was 37.1 (8.76) years. The treatment groups were balanced for these baseline demographic characteristics.

3.1.1.4.3 Baseline disease characteristics

Baseline MS disease characteristics were consistent with a RRMS patient population and were balanced across the treatment groups (Table 2). The median duration of MS since first symptoms was 6.7 years (range 0 to 37 years). The median number of relapses was 2.0 (range 1 to 11) in the previous two years and 1.0 (range 0 to 6) in the previous year. The median baseline EDSS score was 2, identical in all treatment groups (range 0 to 5.5).

Table 2 Clinical MS baseline characteristics – Study D2301 (Source: Table 11-3 of sponsor's Study Report)

	FTY720 1.25mg N=429	FTY720 0.5mg N=425	Placebo N=418	Total N=1272
Duration of MS since first symptom, years				
n	429	425	418	1272
Mean (SD)	8.4 (6.86)	8.0 (6.60)	8.1 (6.35)	8.2 (6.60)
Median	6.9	6.6	7.0	6.7
Range	0 - 37	0 - 35	0 - 32	0 - 37
Number of relapses in the last year				
n	429	425	418	1272
Mean (SD)	1.5 (0.81)	1.5 (0.76)	1.4 (0.73)	1.5 (0.77)
Median	1.0	1.0	1.0	1.0
Range	0 - 6	0 - 5	0 - 6	0 - 6
Number of relapses in the last 2 years				
n	429	424	418	1271
Mean (SD)	2.1 (1.25)	2.1 (1.13)	2.2 (1.19)	2.1 (1.19)
Median	2.0	2.0	2.0	2.0
Range	1 - 10	1 - 11	1 - 10	1 - 11
EDSS				
n	429	425	418	1272
Mean (SD)	2.41 (1.36)	2.30 (1.29)	2.49 (1.29)	2.40 (1.32)
Median	2.00	2.00	2.00	2.00
Range	0.0 - 5.5	0.0 - 5.5	0.0 - 5.5	0.0 - 5.5
MSFC z-score				
n	424	422	413	n/a
Mean (SD)	-0.02 (0.75)	0.06 (0.60)	-0.04 (0.76)	n/a
Median	0.13	0.13	0.09	n/a
Range	-5.9 – 1.3	-2.9 – 1.6	-6.4 – 1.9	n/a

Baseline MRI measures for the FTY720 1.25 mg group were worse than the ones of the other 2 groups. Approximately 40% of the patients showed active lesions on MRI (Table 3).

Table 3 MRI baseline characteristics – Study D2301 (Source: Table 11-4 of sponsor's Study Report)

	FTY720 1.25mg N=429	FTY720 0.5mg N=425	Placebo N=418	Total N=1272
Percentage of patients free of Gd-enhancing T1 lesions - n (%)				
n	424	424	416	1264
	257 (60.6)	263 (62.0)	262 (63.0)	782 (61.9)
Number of Gd enhancing T1 lesions				
n	424	424	416	1264
Mean (SD)	1.8 (4.66)	1.6 (5.57)	1.3 (2.93)	1.6 (4.53)
Median	0.0	0.0	0.0	0.0
Range	0 - 50	0 - 84	0 - 26	0 - 84
Volume of Gd-enhancing T1 lesions (mm^3)				
n	424	424	416	1264
Mean (SD)	197.14 (603.74)	169.87 (601.42)	162.33 (421.21)	176.54 (549.31)
Median	0.00	0.00	0.00	0.00
Range	0.0 - 6852.7	0.0 - 6849.8	0.0 - 2970.0	0.0 - 6852.7
Total volume of T2 lesions (mm^3)				
n	425	424	416	1265
Mean (SD)	6828.70 (8490.54)	6127.71 (7622.97)	6162.40 (7084.84)	6374.63 (7759.71)
Median	3556.50	3303.35	3416.25	3453.30
Range	0.0 - 47734.1	0.0 - 47147.6	0.0 - 37147.8	0.0 - 47734.1
Total volume of T1 hypointense lesions (mm^3)				
n	424	424	416	1264
Mean (SD)	2113.52 (3219.65)	1897.62 (2854.06)	1962.00 (3131.13)	1991.23 (3070.76)
Median	859.55	814.05	811.15	826.90
Range	0.0 - 25885.9	0.0 - 22377.8	0.0 - 20955.9	0.0 - 25885.9
Normalized brain volume (cc)				
n	423	424	414	1261
Mean (SD)	1510.51 (85.94)	1520.84 (83.16)	1512.16 (85.49)	1514.53 (84.92)
Median	1514.69	1528.50	1514.84	1520.22
Range	1217.1 - 1763.8	1143.7 - 1733.7	1229.8 - 1722.6	1143.7 - 1763.8

3.1.1.4.4 Medications taken prior to the start of study drug treatment

A summary of MS disease-modifying drugs (excluding symptomatic treatments) used at any time prior to the start of study drug treatment is presented by treatment group in Table 4. Slightly more than half of all patients were treatment-naïve (approximately 57-60% across the treatment groups). Of those who had been previously treated, interferon had been used most often (367/520 or 70.6%).

Table 4 Prior use of MS disease-modifying drug – Study D2301 (Source: Table 11-5 of sponsor's Study Report)

	FTY720 1.25mg N=429 n (%)	FTY720 0.5mg N=425 n (%)	Placebo N=418 n (%)	Total N=1272 n (%)
Treatment-naïve patients*	259 (60.4)	244 (57.4)	249 (59.6)	752 (59.1)
Previously treated patients	170 (39.6)	181 (42.6)	169 (40.4)	520 (40.9)
Any Interferon beta	125 (29.1)	127 (29.9)	115 (27.5)	367 (28.9)
Interferon beta 1a i.m.	50 (11.7)	65 (15.3)	60 (14.4)	175 (13.8)
Interferon beta 1a s.c.	53 (12.4)	56 (13.2)	49 (11.7)	158 (12.4)
Interferon beta 1b s.c.	44 (10.3)	41 (9.6)	44 (10.5)	129 (10.1)
Glatiramer acetate	52 (12.1)	42 (9.9)	44 (10.5)	138 (10.8)
Natalizumab	1 (0.2)	4 (0.9)	2 (0.5)	7 (0.6)
Other MS medications	43 (10.0)	46 (10.8)	52 (12.4)	141 (11.1)

3.1.1.5 Efficacy Results

The efficacy results presented in this section represent the ones reported by the sponsor and confirmed by the reviewer as well as the results from additional analyses performed by the reviewer.

3.1.1.5.1 Efficacy Results of the Primary Endpoint

The primary analysis of ARR included 1271 patients. One patient in the FTY720 0.5 mg group did not have prior 2-year relapse number, and was not included in the primary analysis. The patient was included in the calculation of unadjusted ARR and other analyses that did not require baseline number of relapses.

The primary analysis of ARR with negative binomial model was performed, and the results reported by the sponsor were confirmed. Analysis of all relapses during the study and confirmed relapses while on treatment were performed using the same model. The following table presents the results from these analyses.

Table 5 Results from analysis of ARR – Study D2301 (Source: reviewer's analysis)

Annualized Relapse Rate (ARR)	FTY720 1.25 mg N=429	FTY720 0.5 mg N=425	Placebo N=418
Confirmed relapses during Study			
Unadjusted (observed)	0.19	0.21	0.47
Adjusted (estimated from model)	0.16	0.18	0.40
95% CI	(0.13, 0.19)	(0.15, 0.22)	(0.34, 0.47)
p-value	<.001	<.001	
Hazard ratio from Cox model	0.38	0.48	
% free of confirmed relapse	75.52	71.06	47.85
Confirmed relapses on Treatment			
Unadjusted (observed)	0.16	0.21	0.48
Adjusted (estimated from model)	0.14	0.18	0.43
p-value	<.001	<.001	
All Relapses during Study			
Unadjusted (observed)	0.26	0.31	0.65
Adjusted (estimated from model)	0.24	0.29	0.62
p-value	<.001	<.001	
Relapse rate at patient level (mean)			
Confirmed relapses on study	0.24	0.23	0.56
Confirmed relapses on Treatment	0.30	0.36	1.12
All relapses	0.32	0.32	0.77

Treatment with both FTY720 1.25 mg and FTY720 0.5 mg resulted in lower aggregate ARRs compared to treatment with placebo, with ARR estimates of 0.16 and 0.18 vs. 0.40, respectively. This corresponded to reductions of 60% and 54% in ARR estimates, for the 1.25 mg and 0.5 mg doses, respectively, which were statistically significant relative to placebo ($p<0.001$ for both comparisons). The difference between the two FTY720 dose groups was not statistically significant ($p=0.238$) from the primary analysis.

Results from analyses of ARR for all relapses, confirmed and non-confirmed, and ARR while patients were on treatment are consistent with the results of confirmed relapse (Table 5) in terms of between-group treatment difference. Results from analysis of ARR on per-protocol (PP) population were similar.

The difference between the adjusted and unadjusted relapse rate is largely contributed by countries of Hungary and Slovakia, which were pooled with Estonia and had a total of 30 patients. Without this pooled country, the estimates of ARR would be 0.17, 0.20, and 0.44 for FTY720 1.25 mg, 0.5 mg, and placebo group, respectively. More details of subgroup analysis can be found in Section 4.

The relapse rate at patient level was higher than the relapse rate at group level in all forms of relapses and in all treatment groups. This was because most patients had 0 relapse, and their

relapse rate was 0 regardless how long they had stayed in the study. Most of these patients stayed in the study until completion.

Other Analysis Related to the Relapse Rate

Time to first confirmed relapse is plotted in the following graph. Treatment difference in time to first confirmed relapse was analyzed using a log-rank test. The difference between each of the FTY720 dose groups and placebo group yielded a nominal p-value of less than 0.001. Median time to first confirmed relapse could not be obtained because more than half of the patients did not have confirmed relapse at study completion or early withdrawal. Hazard ratio, which measures the relative risk of having a relapse, was estimated from the Cox proportional hazard model, included terms of treatment, country, baseline number of relapses and baseline EDSS scores. The estimated hazard ratio for FTY720 1.25 mg and 0.5 mg group relative to the placebo group was 0.38 and 0.48, respectively, with nominal p-values of less than .001 for comparisons of each of the FTY720 dose groups versus placebo.

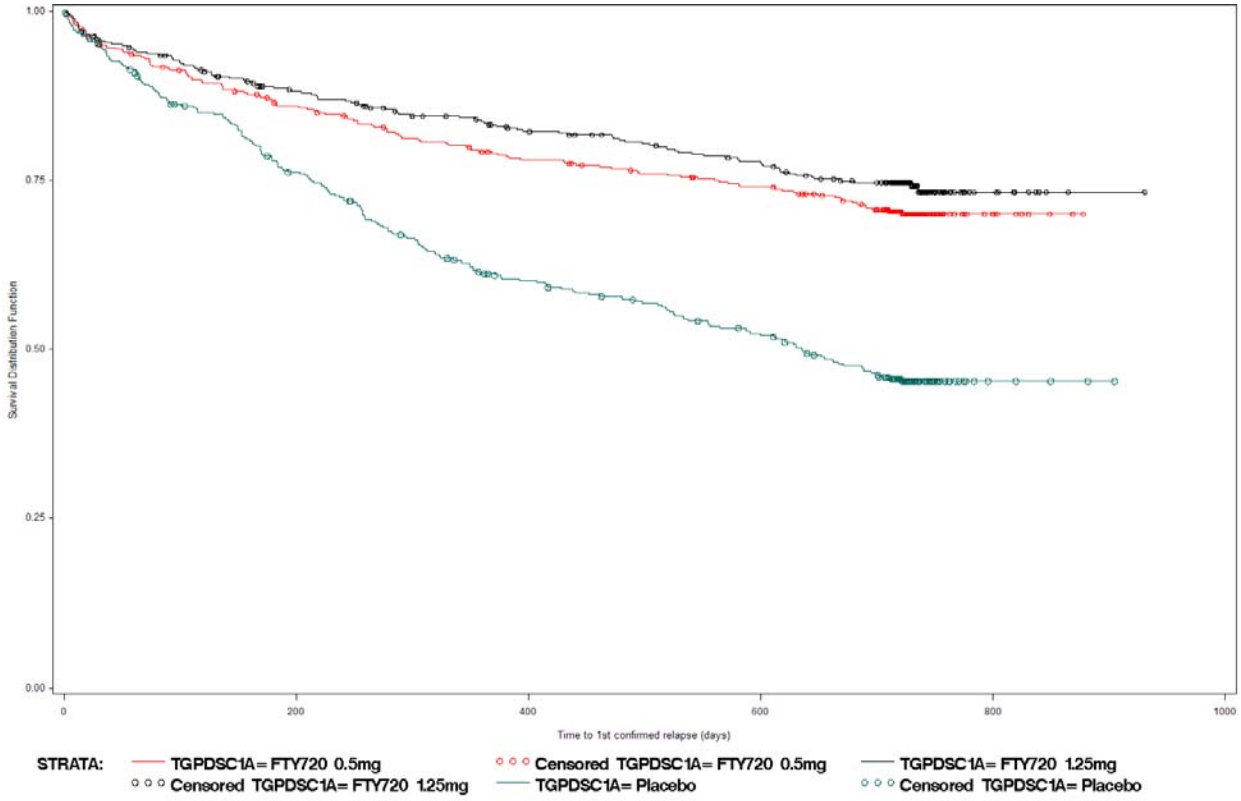


Figure 1 Time to first comfirmed relapse - ITT population (Study D2301) (Source: reviewer's analysis)

3.1.1.5.2 Efficacy Results of Secondary Endpoints

The key secondary efficacy endpoint was time to 3-month confirmed disability progression up to Month 24. Time to disability progression curves for each treatment group were generated by the Kaplan–Meier method and compared by means of the log-rank test.

FTY720 at doses of 1.25 mg and 0.5 mg significantly delayed the time to 3-month confirmed disability progression compared to placebo in the ITT population (log-rank test; p=0.012 and p=0.026, respectively) (Figure 2). The two FTY720 dose groups were not statistically significantly different (p=0.7427). Results from analysis of time to 6-month confirmed disability were similar, with nominal p-values of 0.0044 and 0.0112 for the comparisons of FTY720 1.25 mg and 0.5 mg versus placebo, respectively.

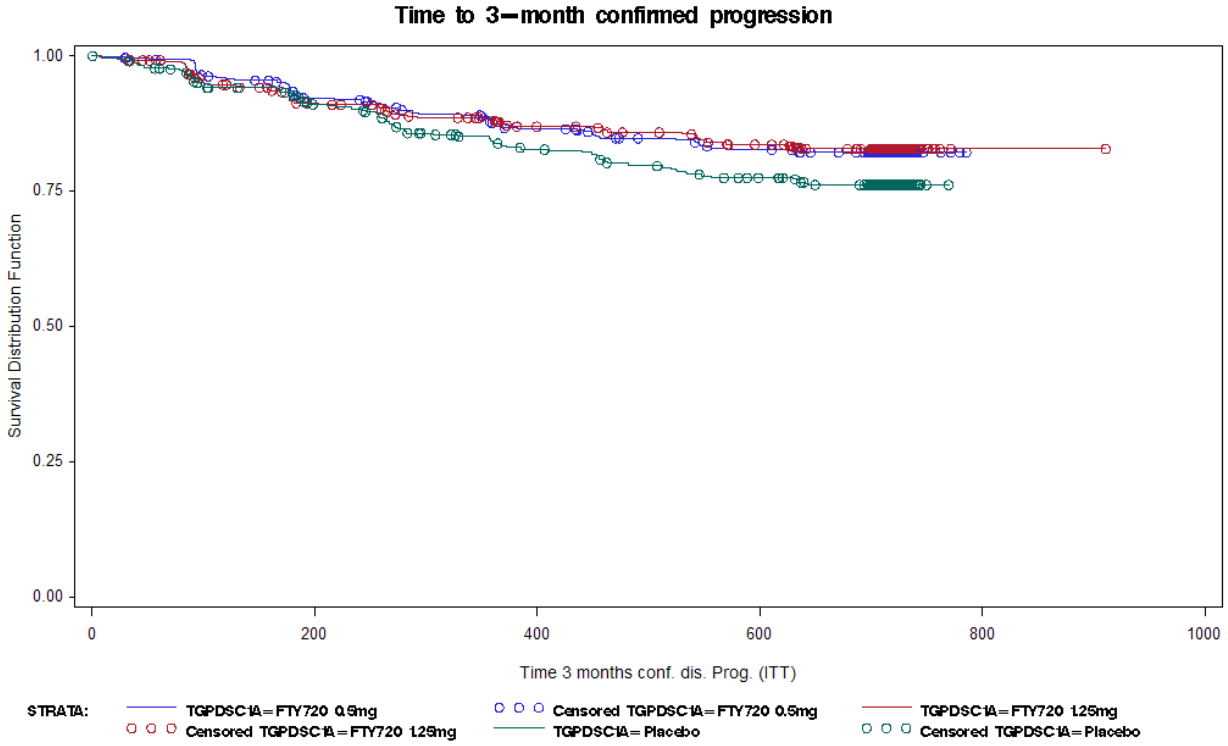


Figure 2 Time to 3-month confirmed disability progression -ITT population (Study D2301) (Source: reviewer's analysis)

The median time to 3-month disability progression could not be estimated because more than 50% of patients in each treatment group were censored. The means of time to disability progression are therefore underestimated due to high censoring, and are not reported here.

The percentage of patients without 3-month confirmed disability progression at Month 24 was higher in both FTY720 treatment groups (84.62% and 83.06% for 1.25 mg and 0.5 mg, respectively) compared with placebo (77.51%). The pairwise comparisons using chi-square test

yielded nominal p-values of 0.008 and 0.043 for FTY720 1.25 mg and 0.5 mg versus placebo, respectively.

Analysis of change from baseline in EDSS scores was performed for ITT patient population. After excluding the EDSS assessments that were performed during a relapse or assessments from unscheduled visit for safety reasons, 1254 had at least one valid EDSS score during the study. Post-hoc analyses using an ANOVA model adjusted for baseline EDSS score and country were performed. Note that some patients discontinued treatment but stayed in the study, and their last assessment of EDSS score represented the one when they were off study drug. Patients who discontinued study drug but stayed in the study were allowed to take alternative MS disease modifying drug. Therefore, analysis of change in EDSS from baseline to last on-treatment score was also performed. The following table presents the results.

Table 6 Change from baseline (LOCF) in EDSS score – Study D2301 (Source: reviewer's analysis)

	FTY720 1.25 mg	FTY720 0.5 mg	Placebo
During Study, mean (SD)			
N	420	422	411
Baseline	2.410 (1.358)	2.302 (1.287)	2.468 (1.271)
Change	0.007 (0.888)	0.002 (0.887)	0.135 (0.957)
Nominal p-value	0.0282	0.0144	
On treatment, mean (SD)			
N	387	411	397
Baseline	2.413 (1.343)	2.293 (1.283)	2.448 (1.272)
Change	-0.031 (0.852)	-0.036 (0.859)	0.092 (0.910)
Nominal p-value	0.0369	0.0126	

It appeared that EDSS scores were little changed during the study or while patients were on treatment. However, the nominal p-values without multiplicity adjustment were below 0.05 for all comparisons of FTY720 dose groups versus placebo, which had a slight increase in EDSS scores.

3.1.2 Study D2302

3.1.2.1 Description of the Study

The primary objective of the study was to compare two doses of FTY720 (1.25 mg and 0.5 mg) with IFN β-1a i.m. to demonstrate that at least 1.25 mg FTY720 was superior to IFN β-1a in terms of annualized relapse rate (ARR) in patients with RRMS treated for up to 12 months.

Key secondary objectives were to demonstrate superiority of FTY720 1.25 mg and 0.5 mg over IFN β-1a in patients with RRMS treated for up to 12 months with respect to: 1) the effect on inflammatory disease activity as measured by the number new/ newly enlarged T2 lesions; and 2) the effect on disability progression as measured by the time to 3-month confirmed disability progression as measured by EDSS.

This was a 12-month, randomized, multicenter, double-blind, double-dummy active-controlled, parallel-group study in patients with RRMS. Patients were randomized to receive an oral fixed dose of FTY720 0.5 mg/day or 1.25 mg/day, or IFN β-1a i.m. 30 μg/week i.m. in a double dummy design.

The study consisted of three phases: a pre-randomization phase lasting for up to 45 days, a 12-month double-blind treatment phase, and an optional extension phase, which is expected to last until FTY720 is commercially available or development is stopped.

The study was conducted in 172 centers in 18 countries, including US. A total of 1275 patients were planned and 1292 patients were actually randomized. The study enrolled patients who were treatment-naïve or previously treated, had diagnosis of MS by 2005 revised McDonald criteria with a relapsing-remitting course, had at least one documented relapse during the previous year or two documented relapses during the previous 2 years prior to randomization, and had Expanded Disability Status Scale (EDSS) score of 0 to 5.5 inclusive.

3.1.2.2 Efficacy Variables

3.1.2.2.1 Primary Efficacy Variable

The primary endpoint was the ARR, which was defined as the number of relapses in a year. Only confirmed relapses were considered for the primary analyses. Refer to Section 3.1.1.2 for definition of confirmed relapse and calculation of relapse rate.

3.1.2.2.2 Key Secondary Variables

There were two key secondary efficacy variables: number of new or newly enlarged T2 lesions on MRI scan at Month 12 and time to 3-month confirmed disability progression at Month 12.

MRI was to be performed at screening, Month 12, and follow-up visit 3 months after the discontinuation of study drug.

Refer to Section 3.1.1.2 Efficacy Variables for definition of 3-month confirmed disability progression.

3.1.2.3 Statistical Analysis Methods

Efficacy analyses for the primary and secondary efficacy endpoints were to be applied to the intent-to-treat population (ITT), which was defined as all patients who were randomized and received at least one dose of study medication. Patients were grouped according to the assigned treatment.

3.1.2.3.1 Analysis of Primary Efficacy Variable

The primary null hypotheses to be tested were: 1) there is no difference in the ARRs between patients treated with the FTY720 1.25 mg and IFN β-1a, and 2) there is no difference in the ARRs between patients treated with the FTY720 0.5 mg and IFN β-1a.

The test of the hypotheses was to be based on a negative binomial regression model for the aggregate ARR adjusting for treatment group, country, baseline number of relapses in previous 2 years, and baseline EDSS as covariates. For the negative binomial regression, the response variable was the number of relapses for each patient. Log of time on study in years was to be used as the offset variable to account for the varying lengths of patients' time in the study. The ARR and its 95% confidence interval (CI) for each treatment group were to be estimated from the model.

For patients who prematurely discontinued study drug, the intent-to-treat approach was to use all relapse data, i.e. relapse data collected after the study drug discontinuation were included in the analyses.

3.1.2.3.2 Analysis of Key Secondary Efficacy Variables

The first key secondary efficacy endpoint was the number of new or newly enlarged T2 lesions at Month 12. Between-treatment comparisons of FTY720 with IFN β-1a were to be performed using a negative binomial model adjusting for treatment group, country, baseline number of relapses in the previous 2 years, and baseline EDSS.

The second key secondary efficacy endpoint was the time to 3-month confirmed disability progression as measured by EDSS during 12 months.

Time-to-event curves for each treatment group were to be generated by the Kaplan–Meier method and compared by means of the log-rank test.

Cox proportional hazard model was a secondary analysis for the time to 3-month confirmed disability progression. The model was adjusted for treatment, country, baseline EDSS and age. Hazard ratios and p-values were to be obtained.

If a patient died due to MS after the start of tentative progression, then the time to disability progression was to be calculated using the onset date of progression. If a patient died due to MS before having progression, then the time to disability progression was to be censored using the date of death.

A patient was to be censored if follow-up ended before a confirmed progression occurred. The disability progression occurred after the 9 month visit could not be confirmed due to the 12-month study duration. Hence, they were to be treated as the censoring data in the analysis.

3.1.2.3.3 Multiplicity Adjustment

To control the overall type-I error rate of the study, a multiplicity adjustment was to be applied to the primary and key secondary endpoints. There was one primary endpoint and two key secondary endpoints with two doses, which yielded six FTY720 (1.25 mg and 0.5 mg) comparisons vs. IFN β-1a. The testing of FTY720 comparisons vs. IFN β-1a was to be done in a hierarchical order according to as follows:

1. FTY720 1.25 mg, ARR
2. FTY720 0.5 mg, ARR
3. FTY720 1.25 mg, the number of new and newly enlarged T2 lesions at 12 months
4. FTY720 0.5 mg, the number of new and newly enlarged T2 lesions at 12 months
5. FTY720 1.25 mg, disability progression
6. FTY720 0.5 mg, disability progression.

Each testing was to be performed at a significant level of 0.05 for these six comparisons. However, the lower-rank testing was to be performed only when every high-rank testing was statistically significant.

3.1.2.4 Study Patients

3.1.2.4.1 Disposition of patients

A total of 1573 patients were screened for participation in this study. Of the 1292 patients who were randomized, 1153 (89.2%) completed the study (86.6% in the FTY720 1.25 mg group, 92.3% in the FTY720 0.5 mg group, and 88.7% in the IFN β-1a group). A total of 1123 patients (86.9%) completed the study on study drug (84.0% in the FTY720 1.25 mg group, 89.3% in the FTY720 0.5 mg group, and 87.4% in the IFN β-1a group).

The most common reason for discontinuation of study drug overall was AEs (4.6% of all patients, 7.5% for FTY720 1.25 mg, 3.7% for FTY720 0.5 mg, and 2.8% for IFN β-1a), followed by withdrawal of consent (2.7% of all patients; 2.3% for FTY720 1.25 mg, 2.1% for FTY720 0.5 mg, and 3.7% for IFN β-1a). Of the 157 patients who discontinued study drug, 30 patients remained in the study and completed the abbreviated schedule of assessments through the Month 12 visit.
Table 7 presents the disposition of patients.

Table 7 Patient disposition – Study D2302 (Source: Table 10-2 of sponsor's Study Report)

	FTY720 1.25mg N=426 n (%)	FTY720 0.5mg N=431 n (%)	Interferon beta-1a i.m. N=435 n (%)	Total N=1292 n (%)
Completed study	369 (86.6)	398 (92.3)	386 (88.7)	1153 (89.2)
On study drug*	358 (84.0)	385 (89.3)	380 (87.4)	1123 (86.9)
Off study drug**	11 (2.6)	13 (3.0)	6 (1.4)	30 (2.3)
Discontinued from the study	57 (13.4)	33 (7.7)	49 (11.3)	139 (10.8)
Adverse event(s)	26 (6.1)	9 (2.1)	9 (2.1)	44 (3.4)
Subject withdrew consent	11 (2.6)	9 (2.1)	16 (3.7)	36 (2.8)
Administrative problems	6 (1.4)	2 (0.5)	7 (1.6)	15 (1.2)
Unsatisfactory therapeutic effect	3 (0.7)	3 (0.7)	7 (1.6)	13 (1.0)
Abnormal laboratory value(s)	4 (0.9)	6 (1.4)	1 (0.2)	11 (0.9)
Abnormal test procedure result(s)	4 (0.9)	3 (0.7)	3 (0.7)	10 (0.8)
Lost to follow-up	1 (0.2)	1 (0.2)	4 (0.9)	6 (0.5)
Death	2 (0.5)	0	0	2 (0.2)
Protocol violation	0	0	2 (0.5)	2 (0.2)
Discontinued study drug	62 (14.6)	44 (10.2)	51 (11.7)	157 (12.2)
Adverse event(s)	32 (7.5)	16 (3.7)	12 (2.8)	60 (4.6)
Subject withdrew consent	10 (2.3)	9 (2.1)	16 (3.7)	35 (2.7)
Abnormal laboratory value(s)	8 (1.9)	7 (1.6)	3 (0.7)	18 (1.4)
Unsatisfactory therapeutic effect	5 (1.2)	5 (1.2)	7 (1.6)	17 (1.3)
Abnormal test procedure result(s)	3 (0.7)	4 (0.9)	4 (0.9)	11 (0.9)
Administrative problems	1 (0.2)	2 (0.5)	3 (0.7)	6 (0.5)
Lost to follow-up	1 (0.2)	0	4 (0.9)	5 (0.4)
Protocol violation	0	1 (0.2)	2 (0.5)	3 (0.2)
Death	1 (0.2)	0	0	1 (0.1)
Subject's condition no longer requires study drug	1 (0.2)	0	0	1 (0.1)

* 'On study drug': Patients who took study drug until the study completion.
** 'Off study drug': Patients who completed the study but discontinued study drug.
Note: The total number of patients who discontinued the study includes 12 patients who were randomized in error and never received study drug.
Note: This table displays the number of patients with the primary reason for discontinuation recorded as "adverse event;" tables in Section 12 display patients with AEs with outcome of "discontinuation of study drug" and therefore numbers can be expected to differ.
Note: 2 additional patients in the FTY720 1.25 mg treatment group died after data base lock: patients PID 254/00011 and PID 331/00011, further discussed in Section 12.3. The events occurred 3 and 6 months after data base lock, respectively, and therefore do not appear in the tables and listings but do appear in the patient narratives.
Source: PT-Tables 14.1-1.1 and 14.1-1.2

The ITT patient population included 1280 subjects. Overall, 12 patients were excluded from both ITT and safety population because they were randomized in error and did not receive study drug. The PP population included all ITT patients who did not have any major protocol deviations, and 34 subjects were excluded from the PP population. The three most common protocol deviations which excluded patients from the PP population were: 1) patient took the wrong treatment medication for less than 3 months (i.e., wrong randomization was inadvertently dispensed to the

patient at one of the study visits by the study staff) (0.9% of all patients); 2) unblinding through MRI (i.e., MRI information was inadvertently shared between the local neuroradiologist and investigator) (0.8% of all patients); and 3) not following per protocol blinding procedures (PI accidentally saw hematology results) (0.7% of all patients).

3.1.2.4.2 Baseline Demographic Characteristics

The groups were balanced for age, sex, and race. Approximately two-thirds of patients were female (67.3% female vs. 32.7% male) and the majority (94.1%) of patients in all groups were Caucasian. The median age was 36 years.

3.1.4.2.3 Baseline Disease Characteristics

Across all treatment groups the mean duration of MS since first symptoms was 7.4 years (median 5.9 years) with an average of 2.2 relapses in the previous 2 years, 1.5 relapses in the previous year, and a mean baseline EDSS score of 2.21. Overall, the groups were balanced for all MS disease baseline characteristics. However, the proportion of patients with EDSS 5.5 at baseline was highest in the FTY720 1.25 mg group (14/420, 3.33%) compared to 11/429 (2.56%) for the FTY720 0.5 mg group and 6/431 (1.39%) for the IFN β-1a group. The MS disease characteristics of patients at baseline are summarized by treatment group in

Table 8.

Appears This Way On Original

Table 8 Clinical MS baseline characteristics - Study D2302 (Source: Table 11-3 of sponsor's Study Report)

	FTY720 1.25mg N=426	FTY720 0.5mg N=431	Interferon beta-1a i.m. N=435	Total N=1292
Duration of MS since first symptom (years)				
n	420	429	431	1280
Mean (SD)	7.3 (5.96)	7.5 (6.20)	7.4 (6.33)	7.4 (6.16)
Median	6.0	5.8	5.8	5.9
Range	0 - 33	0 - 34	0 – 40*	0 - 40*
Number of relapses in the last year				
n	425	431	435	1291
Mean (SD)	1.5 (0.87)	1.5 (1.19)	1.5 (0.79)	1.5 (0.97)
Median	1.0	1.0	1.0	1.0
Range	0 - 7	0 - 20*	0 - 6	0 – 20*
Number of relapses in the last 2 years				
n	425	431	434	1290
Mean (SD)	2.2 (1.19)	2.3 (2.20)	2.3 (1.22)	2.2 (1.61)
Median	2.0	2.0	2.0	2.0
Range	1 - 8	1 – 40*	1 - 12	1 - 40
EDSS				
n	420	429	431	1280
Mean (SD)	2.21 (1.311)	2.24 (1.326)	2.19 (1.261)	2.21 (1.299)
Median	2.00	2.00	2.00	2.00
Range	0.0 - 5.5	0.0 - 5.5	0.0 - 5.5	0.0 - 5.5
MSFC z-score				
n	416	424	423	1263
Mean (SD)	-0.006 (0.7272)	0.007 (0.6327)	0.005 (0.6159)	0.002 (0.6595)
Median	0.106	0.159	0.128	0.124
Range	-5.35 – 2.04	-5.23 – 1.19	-2.81 – 2.51	-5.35 – 2.51
MSFC subscale: 25-foot timed walking test (seconds)				
n	420	427	428	1275
Mean (SD)	7.20 (10.690)	6.71 (7.499)	6.47 (5.736)	6.79 (8.216)
Median	5.00	5.15	5.00	5.05
Range	2.9 - 126.0	2.3 - 121.0	2.7 - 55.0	2.3 - 126.0
MSFC subscale: 9-hole peg test (seconds)				
n	420	426	428	1274
Mean (SD)	22.58 (14.344)	22.34 (10.091)	21.98 (7.992)	22.30 (11.100)
Median	20.10	20.03	20.00	20.04
Range	8.8 - 196.8	11.0 - 120.5	4.8 - 101.0	4.8 - 196.8
MSFC subscale: PASAT-3 (number of correct answers)				
n	416	424	424	1264
Mean (SD)	47.9 (11.15)	48.3 (11.09)	47.7 (11.94)	48.0 (11.39)
Median	51.0	51.0	52.0	52.0
Range	2 - 60	0 - 60	0 - 60	0 - 60

The mean number and the volume of Gd-enhanced T1-weighted lesions at baseline was higher in the FTY720 1.25 mg group (1.5 and 147.5, respectively) than in the FTY720 0.5 mg group (1.0 and 93.9, respectively) and the IFN β-1a group (1.1 and 100.7, respectively), and the difference (vs. the IFN β-1a group) carried a p-value of 0.068. The total volume of T2 lesions as well as all other baseline MRI characteristics was comparable among treatment groups. MRI characteristics for patients at baseline are summarized by treatment group in Table 9.

Table 9 MRI baseline characteristics - Study D2302 (Source: Table 11-4 of sponsor's Study Report)

	FTY720 1.25mg N=426	FTY720 0.5mg N=431	Interferon beta-1a i.m. N=435	Total N=1292
Proportion of patients free of Gd-enhanced T1 lesions n (%)				
n	412	427	425	1264
	270 (65.5)	288 (67.4)	268 (63.1)	826 (65.3)
Number of Gd-enhanced T1 lesions				
n	412	427	425	1264
Mean (SD)	1.5 (4.77)	1.0 (2.81)	1.1 (2.80)	1.2 (3.57)
Median	0.00	0.00	0.00	0.00
Range	0.0 - 66	0.0 - 29	0.0 - 36	0.0 - 66
Volume of Gd-enhanced T1 lesions (mm^3)				
n	412	427	425	1264
Mean (SD)	147.5 (667.21)	93.9 (288.05)	100.7 (263.55)	113.7 (443.54)
Median	0.0	0.0	0.0	0.0
Range	0 - 11507	0 - 3250	0 - 2609	0 - 11507
Total volume of T2 lesions (mm^3)				
n	413	428	425	1266
Mean (SD)	5085.4 (5962.05)	5169.6 (6641.97)	4923.6 (5710.90)	5059.5 (6116.41)
Median	3095.9	2381.8	2901.1	2786.6
Range	0 - 38870	0 - 46280	0 - 38712	0 - 46280
Total volume of T1 hypointense lesions (mm^3)				
n	413	428	425	1266
Mean (SD)	1386.7 (2298.52)	1620.4 (3107.07)	1404.2 (2357.82)	1471.6 (2618.03)
Median	454.9	444.9	420.6	439.2
Range	0 - 20399	0 - 30610	0 - 19561	0 - 30610
Normalized brain volume (cc)				
n	409	421	420	1250
Mean (SD)	1526.2 (76.37)	1524.1 (83.88)	1526.7 (77.93)	1525.7 (79.43)
Median	1527.8	1526.2	1533.3	1529.5
Range	1300 - 1794	1185 - 1862	1231 - 1762	1185 - 1862

About 40-45% of the patients were treatment-naïve. Of the 732 patients who were previously treated with at least one MS disease-modifying drug, 552 patients were still receiving an MS disease-modifying drug within the 3 months prior to the start of study drug treatment. Approximately one third of these patients had received treatment with 2 or more MS disease-modifying drugs. Patients who were receiving MS medications prior to the start of study drug

treatment were allowed to enter the study without a washout period. Table 10 presents the information of prior use of MS disease-modifying drug.

Table 10 Prior use of MS disease-modifying drug - Study D2302 (Source: Table 11-5 of sponsor's Study Report)

	FTY720 1.25mg (N=426) n (%)	FTY720 0.5mg (N=431) n (%)	Interferon beta-1a i.m. (N=435) n (%)	Total (N=1292) n (%)
Treatment-naïve patients*	177 (41.5)	193 (44.8)	190 (43.7)	560 (43.3)
Previously treated patients	249 (58.5)	238 (55.2)	245 (56.3)	732 (56.7)
Any interferon beta	209 (49.1)	219 (50.8)	207 (47.6)	635 (49.1)
Interferon beta 1a i.m.	118 (27.7)	119 (27.6)	118 (27.1)	355 (27.5)
Interferon beta 1a s.c.	79 (18.5)	89 (20.6)	72 (16.6)	240 (18.6)
Interferon beta 1b s.c.	57 (13.4)	59 (13.7)	69 (15.9)	185 (14.3)
Glatiramer acetate	67 (15.7)	57 (13.2)	67 (15.4)	191 (14.8)
Natalizumab	3 (0.7)	4 (0.9)	1 (0.2)	8 (0.6)

* Treatment-naïve patients are defined as those not receiving any of the approved 5 MS disease-modifying drugs listed above (Section 9.7.1.1.3).
Source: PT-Table 14.1-3.10

3.1.2.5 Efficacy Results

3.1.2.5.1 Efficacy Results of the Primary Endpoint

One patient in the IFN β-1a group did not have prior 2-year relapse number, and was not included in the primary analysis. The patient was included in the calculation of unadjusted ARR and other analyses that did not require baseline number of relapses. Another patient in the FTY720 0.5 mg group had 40 relapses during the 2 years prior to study entry, and the patient baseline relapse number was changed to 24, the maximum possible in a 2-year period. This change did not resulted in different estimates of ARR from the ones obtained by the sponsor in the primary analysis. The results for the primary efficacy analysis (aggregate ARR analyzed using negative binomial regression), analysis of all relapses during the study and confirmed relapses while on treatment are shown in the following table.

Table 11 Results from analysis of ARR – Study D2302 – (Source: reviewer's analysis)

ARR	FTY720 1.25 mg N=420	FTY720 0.5 mg N=429	IFN β-1a N=431
Confirmed relapses during Study			
Unadjusted (observed)	.26	.21	.43
Adjusted	.20	.16	.33
95% CI	(.16, .26)	(.12, .21)	(.26, .41)
p-value	<.001	<.0001	
Hazard ratio from Cox model	.63	.52	
% free of confirmed relapse	80.48	82.52	70.07
Confirmed relapses on Treatment			
Unadjusted	.25	.21	.43
Adjusted	.20	.16	.34
p-value	.0002	<.0001	
All Relapses			
Unadjusted	.33	.30	.63
Adjusted	.28	.24	.51
p-value	<.0001	<.0001	
Relapse rate at patient level (mean)			
Confirmed relapses on study	0.26	0.21	0.43
Confirmed relapses on Treatment	0.25	0.21	0.43
All relapses	0.33	0.30	0.63

Treatment with both FTY720 1.25 mg and FTY720 0.5 mg resulted in a significantly lower ARR compared to treatment with IFN β-1a with ARR estimates of 0.20 and 0.16 vs. 0.33, respectively. This corresponded to reductions of 38% and 52% in ARR estimates, respectively, which were statistically significant ($p<0.001$ for both comparisons). The difference between the two FTY720 dose groups was not statistically significant ($p=0.152$) from the primary analysis.

Results from analyses of ARR for all relapses, confirmed and non-confirmed, and ARR while patients were on treatment are consistent with the results from analysis of confirmed relapse.

Three countries had a large effect on estimates of the relapse rate: Korea with 18 patients, Greece with 31 patients and Switzerland with 22 patients. More details of subgroup analysis including difference between US and non-US patient population can be found in Section 4.

Time to first confirmed relapse is plotted in the following graph. Treatment difference in time to first confirmed relapse was analyzed using a log-rank test and hazard ratios were estimated from the Cox proportional hazard model. The difference between each of the FTY720 dose groups and IFN β-1a group yielded a nominal p-value of less than 0.001 from the log-rank test. Median time to first confirmed relapse could not be obtained because more than half of the patients did not have confirmed relapse at study completion or early withdrawal.

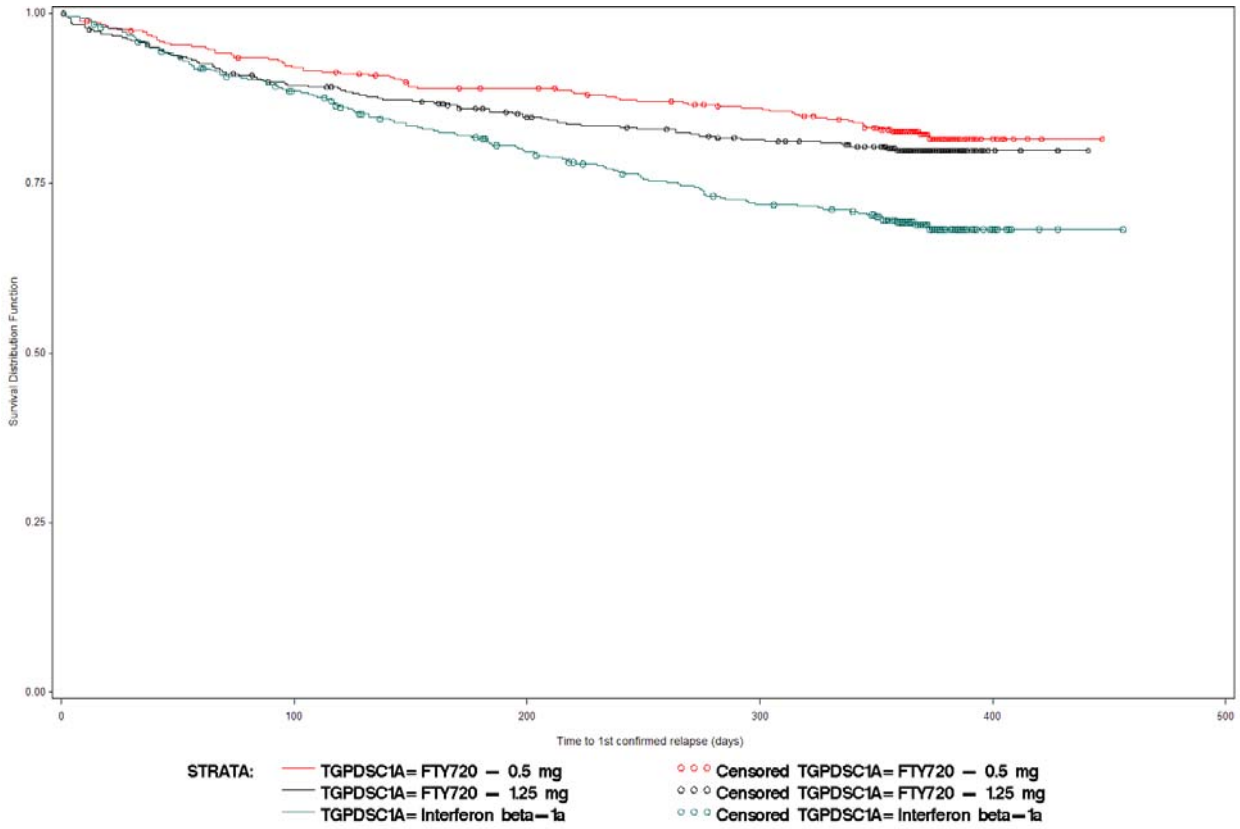


Figure 3 Time to confirmed relapse - ITT population (Study D2302) (Source: reviewer's analysis)

Results from analyses of ARR on PP population and ARR at the patient level were consistent with the results from the primary analysis.

3.1.2.5.2 Efficacy Results of Key Secondary Endpoints

New or newly enlarged T2 lesions at Month 12

The results of MRI measure were from the sponsor. MRI T2 lesions were not read as intended originally. The sponsor later filed an amendment to correct the T2 lesion count. However, results from 18 of the subjects could not be included because correct readings of the subjects were not available. The reviewer will file an addendum of the review when correct reading of the scans of 18 subjects have completed and analyzed. The following MRI results were reported by the sponsor in the original submission before the amendment.

The number of new or newly enlarged T2 lesions at Month 12 was compared between treatment groups using a negative binomial regression model adjusting for the same covariates used in the primary efficacy analysis (treatment, country, baseline number of relapses in the previous 2 years, and baseline EDSS), and the results are shown in Table 12.

Table 12 Mean number of new or newly enlarged T2 lesions at Month 12 – Study D2302 (Source: Table 11-8 of sponsor's Study Report)

	FTY720 1.25mg N=420	FTY720 0.5mg N=429	Interferon beta-1a i.m. N=431
n	356	380	365
Mean (SD)	1.4 (2.51)	1.5 (3.50)	2.1 (4.86)
Median	1.0	0.0	1.0
Range	0 - 22	0 - 32	0 - 60
P-value for treatment comparison of FTY720 vs. Interferon beta-1a i.m. (negative binomial regression with covariates)	0.017*	0.053	–

n=the number of patients with evaluable MRI at baseline and Month 12
P-value is calculated using a negative binomial model adjusting for treatment, country, baseline number of relapses in the previous 2 years, and baseline EDSS.

For the ITT population both the FTY720 1.25 mg and FTY720 0.5 mg treatment groups had a lower mean number of new or newly enlarged T2 lesions at Month 12 compared to the IFN β-1a group, which reached statistical significance for the FTY720 1.25 mg group (p=0.017) and did not reach statistical significance for the FTY720 0.5 mg group (p=0.053). In the PP population, the mean number of new or newly enlarged T2 lesions at Month 12 compared to the IFN β-1a group, did not reach statistical significance for the FTY720 1.25 mg group (p=0.067) or the FTY720 0.5 mg group (p=0.052).

A sensitivity analysis for this efficacy endpoint was performed using the same negative binomial regression model pre-specified in the analysis plan on all available data (two FTY720 arms and IFN β-1a) to fully assess the effect of the covariates on treatment responses. Analysis results showed that both FTY720 1.25 mg and 0.5 mg treatment groups were superior to IFN β-1a (p=0.007 and 0.038).

Time to 3-month confirmed disability progression at Month 12

There was no difference between either of the two FTY720 treatment groups and the IFN β-1a group in the time to 3-month confirmed disability progression based on log-rank test (p-values are 0.4979 and .2475 for FTY720 1.25 mg and 0.5 mg versus IFN β-1a, respectively).

Altogether, 85 patients had confirmed disability progression, 27 in the 1.25 mg FTY720 group, 25 in the 0.5 mg FTY720 group, and 33 in the placebo. The proportion of patients who were free of disability progression were 94.20% and 92.41% for FTY720 1.25 mg and 0.5 mg groups, respectively, compared to 93.66% for the IFN β-1a group. The Kaplan-Meier curve for the time to 3-month confirmed disability progression at Month 12 is shown in Figure 4.

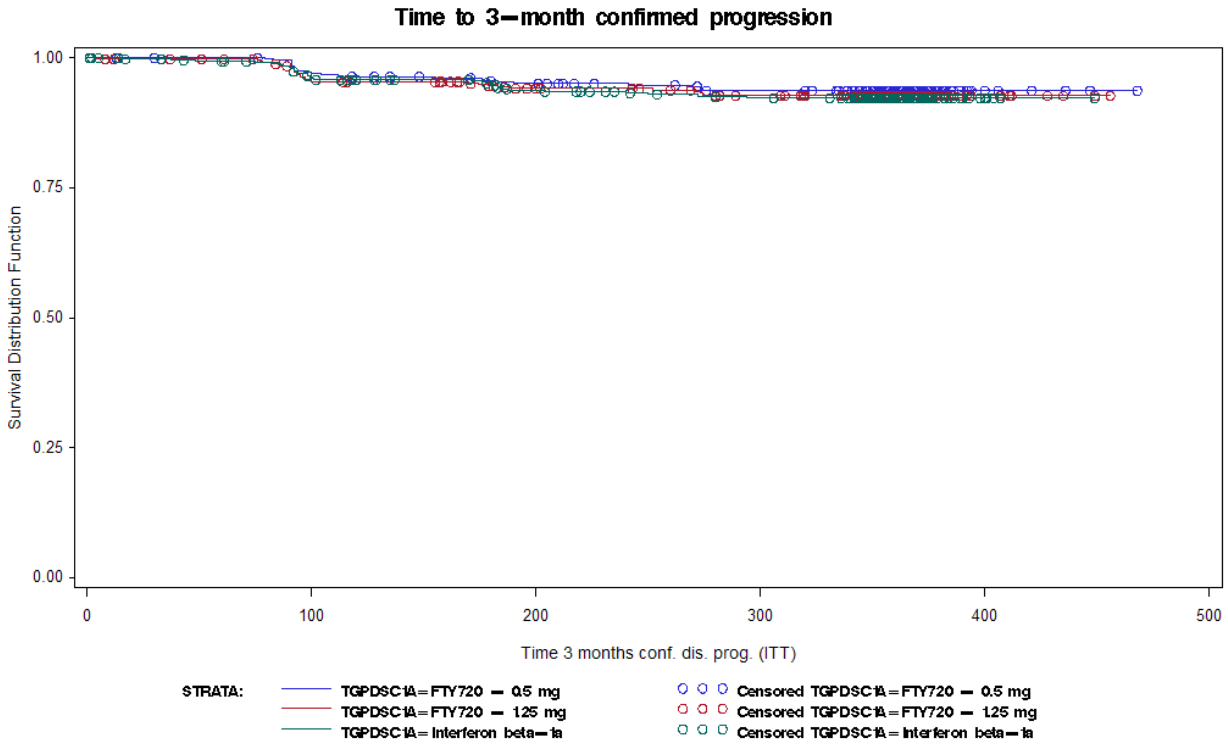


Figure 4 Time to 3-month confirmed disability progression - ITT population (Study D2302) (Source: reviewer's analysis)

In order to examine whether or not the lack of treatment difference in disability progression was due the short length of this study as compared to Study D2301, which was a 2-year study, the first year data of D2301 was analyzed for disability progression and compared to results of D2302. Note that patients who had onset of disability progression after 9 months could not be confirmed in Study D2302. Therefore, the analysis of data in Study D2301 used cutoff time of 290 days, which approximated 9 months plus 14 days of window period. Patients in Study D2301 who did not have onset of disability progression by 290 days were censored.

Table 13 Comparison of disability progression rate for Studies D2301 and D2302 (Source: reviewer's analysis)

D2301 first 9 months	FTY720 1.25 mg N=429	FTY720 0.5 mg N=425	Placebo N=418
Number (%) progressed	44 (10.26%)	43 (10.12%)	59 (14.11%)
p-value	P=0.1024	P=0.0590	
D2302	FTY720 1.25 mg N=420	FTY720 0.5 mg N=429	IFN β-1a N=431
Number (%) progressed	27 (6.43%)	25 (5.83%)	33 (7.66%)
p-value	0.4979	0.2475	

Difference in time to 3-month disability progression for Study D2301 did not reach statistical significance level for either of the FTY720 dose groups but had much smaller p-value compared to Study D2302: p=0.1024 for comparison of FTY720 1.25 mg versus placebo and p=0.0590 for comparison of FTY720 0.5 mg versus placebo. In Study D2302, p-values from comparisons of FTY720 1.25 mg and 0.5 mg versus IFN β-1a were 0.4979 and 0.2475, respectively.

Change from baseline in EDSS scores were performed for ITT patient population. After excluding the EDSS assessments that were performed during a relapse or assessments from unscheduled visit for safety reasons, 1248 patients had at least one valid EDSS score during the study. Post-hoc analyses using an ANOVA model adjusted for baseline EDSS score and country were performed. Note that some patients discontinued treatment but stayed in the study, and for those patients their last assessment of EDSS score represented the one when they were off study drug. Patients who discontinued study drug but stayed in the study were allowed to take alternative MS disease modifying drug. Therefore, analysis of change in EDSS from baseline to last on-treatment score was also performed. The following table presents the results.

Table 14 Change from baseline (LOCF) in EDSS score – Study D2302 (Source: reviewer's analysis)

	FTY720 1.25 mg	**FTY720 0.5 mg**	**IFN β-1a**
During Study, mean (SD)			
N	408	423	417
Baseline	2.212 (1.299)	2.243 (1.334)	2.159 (1.249)
Change	-0.103 (0.862)	-0.084 (0.778)	0.016 (0.798)
Nominal p-value	0.0421	0.1005	
On treatment, mean (SD)			
N	394	412	407
Baseline	2.206 (1.290)	2.239 (1.332)	2.154 (1.257)
Change	-0.126 (0.806)	-0.108 (0.800)	0.010 (0.781)
Nominal p-value	0.0182	0.0504	

There was a small decrease in EDSS scores in both of the FTY720 dose groups and small increase in the IFN β-1a group. The nominal p-values without multiplicity adjustment for the treatment difference in this post-hoc analysis were below 0.05 in the comparison between FTY720 1.25 mg group versus placebo group, but were above 0.05 for the comparison between FTY720 0.5 mg group versus placebo group.

3.2 Evaluation of Safety

Refer to Clinical Review by Dr. Heather Fitter and Safety Review by Dr. Lourdes Villalba for Evaluation of Safety.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Analyses of relapse rate by gender and age group were performed. The majority of patients were Caucasians, and analysis by race was not performed. The following table presents the estimated relapse rate and the p-value for the comparison between each of the FTY720 dose group and the control group in the subgroup population. Due to the large number of countries and relatively small number of patients in each sub-population, accurate estimate of relapse rate from the negative binomial model could not be obtained, and unadjusted relapse rates are presented. In both studies, the data did not suggest gender or age difference in relapse rate.

Table 15 ARR by gender and age group – Study D2301 (Source: reviewer's analysis)

Study D2301	**FTY720 1.25 mg**	**FTY720 0.5 mg**	**Placebo**
Unadjusted relapse rate			
Overall Population			
N	429	425	418
ARR	0.19	0.21	0.47
Sex			
Male, n	134	129	120
ARR	0.21	0.18	0.54
Female, n	295	296	298
ARR	0.18	0.23	0.44
Age			
≤ 37 years, n	204	237	215
ARR	0.19	0.20	0.53
> 37 years, n	225	188	203
ARR	0.19	0.23	0.40

Table 16 ARR by gender and age group - Study D2302 (Source: reviewer's analysis)

Study D2302	**FTY720 1.25 mg**	**FTY720 0.5 mg**	**IFN β-1a**
Unadjusted relapse rate			
Overall Population			
N	420	429	431
ARR	0.26	0.21	0.43
Sex			
Male, n	132	148	139
ARR	0.29	0.21	0.34
Female, n	288	281	292
ARR	0.24	0.21	0.47
Age			
≤ 37 years, n	232	226	244
ARR	0.25	0.18	0.48
> 37 years, n	188	203	187
ARR	0.28	0.24	0.36

4.2 Other Special/Subgroup Populations

Both studies were conducted in a large number of countries globally. Study D2301 was conducted outside of US in 22 countries. A few countries with small number of patients were pooled together to form larger pooled countries: UK and Ireland were pooled, Greece and Israel were pooled, and Estonia, Hungary and Slovakia were pooled. The variation of the pooled countries of Estonia, Hungary and Slovakia was so big that the estimates of ARR and confidence intervals could not be obtained if the pooling was not formed. Among the 3 countries, Hungary and Slovakia belonged to the same region while Estonia was a single country separated from others and had little effect. Altogether, 30 patients were from these 3 countries: 5 in Estonia, 12 in Hungary, and 13 in Slovakia. The following table presents the ARR estimates with the p-values in the overall patient population and in patient population without these 3 countries.

Table 17 Analysis of ARR by region - Study D2301 (Source: reviewer's analysis)

D2301	**FTY720 1.25 mg N=429**	**FTY720 0.5 mg N=425**	**Placebo N=418**
Overall Patient Population			
Adjusted ARR	0.16	0.18	0.40
95% CI	(0.13, 0.19)	(0.15, 0.22)	(0.34, 0.47)
p-value	<.001	<.001	
Excluding patients in 3 countries			
n	417	416	409
Adjusted ARR	0.17	0.20	0.44
95% CI	(0.14, 0.21)	(0.17, 0.24)	(0.37, 0.51)
Nominal p-value	<.001	<.001	

Although these 30 patients constituted only less than 2.5% of the total patient population, the estimates of ARR were quite different without them. The p-values were little changed.

Study D2302 was conducted in 18 US and non-US countries. ARR was estimated by sub-populations of US and non-US patients. Korea was the only country in Asia and Greece was the only country in East Europe. These two countries had small number of patients, 18 and 31, respectively, but had large effect in the estimation of ARR. Therefore, estimates of ARR from non-US countries excluding these two countries are also provided. The results are presented in the following table.

Table 18 Analysis of ARR by region - Study D2302 (Source: reviewer's analysis)

D2302	**FTY720 1.25 mg N=420**	**FTY720 0.5 mg N=429**	**IFN β-1a N=430**
Overall Adjusted ARR	0.20	0.16	0.33
By Region			
US, n	42	42	45
Adjusted ARR	0.16	0.28	0.28
95% CI	(0.075, 0.341)	(0.157, 0.499)	(0.163, 0.496)
Nominal p-value	0.2922	0.9043	
Non-US, n	378	387	386
Adjusted ARR	0.21	0.15	0.33
95% CI	(0.158, 0.271)	(0.110, 0.199)	(0.260, 0.424)
Nominal p-value	<.001	<.0001	
Non-US excluding Korea and Greece, n	360	370	372
Adjusted ARR	0.24	0.17	0.39
95% CI	(0.186, 0.307)	(0.132, 0.229)	(0.310, 0.478)
Nominal p-value	<.001	<.0001	

The number of patients in US sites was about 10% of the total patient population, and the estimates of ARR were quite different from the estimates of the non-US patient population. The estimated ARR for FTY720 0.5 mg group and the placebo group were the same (0.28), and estimated ARR for FTY720 1.25 mg group was much smaller (0.16) than the estimate from the overall patient population (0.20).

Annual relapse rate in subgroups with respect to prior MS-drug use, baseline EDSS score and baseline number of Gd-enhancing lesions were reported by the sponsor. The data suggests that treatment-naïve and less severe patients had lower relapse rate than the more severe patients, which was normally expected (results not shown).

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The two pivotal studies collectively provided sufficient evidence that FTY720 at doses 1.25 mg or 0.5 mg is effective in treating patients with relapsing form of multiple sclerosis. No major statistical issues were identified.

5.2 Conclusions and Recommendations

The efficacy results obtained from the analyses of the two pivotal studies D2301 and D2302 support the conclusion that FTY720 is effective in treating patients with relapsing form of multiple sclerosis.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22527	ORIG-1	NOVARTIS PHARMACEUTICALS CORP	FINGOLIMOD HCL ORAL CAPSULES

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

/s/

XIAORONG YAN
06/29/2010

KUN JIN
06/29/2010
I concur with the review.

KOOROS MAHJOOB
07/01/2010
I concur with this review

Statistical Review and Evaluation

CARCINOGENICITY STUDIES

IND/NDA Number:	NDA 21-527
Drug name:	Fingolimod HCl (Gilenia)
Indication(s):	104 Week Carcinogenicity in Rats and Mice
Applicant:	Novartis Pharmaceuticals Corporation
Documents Reviewed:	Electronic submission Dated 2009-06-15 Electronically submitted dataset Dated 2010-01-20
Review Priority:	Standard
Biometrics Division:	Division of Biometrics 7
Statistical Reviewer:	Matthew Jackson, PhD
Concurring Reviewer:	Karl Lin, PhD
Medical Division:	Division of Neurology Products
Reviewing Pharmacologist:	Richard Siarey, PhD
Project Manager:	Jacqueline Ware
Keywords:	Carcinogenicity Animal

Contents

List of Tables

List of Figures

1 Background

In this submission the sponsor included reports of two animal carcinogenicity studies, one in rats and one in mice. These studies were intended to assess the carcinogenic potential of FTY720 in rats and mice when administered orally by gavage once daily at appropriate drug levels for about 104 weeks. Results of this review have been discussed with the reviewing pharmacologist Dr. Siarey.

In this review, 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.

2 Rat study

Two separate experiments were conducted, one in male and one in female rats. In each of these two experiments there were four treated groups and two identical control groups. Three hundred and six Wistar rats of each sex were randomly allocated to treated and control groups of equal size (i.e. 51 animals). The dose levels for treated groups were 0.05, 0.15, 0.5, and 2.5 mg/kg/day. In this review these dose groups were referred to as the low, medium, high and very high dose groups, respectively. The vehicle for the test was a solution of graded millipore water, in concentrations of 0.0025%, 0.0075%, 0.025% and 0.125% for the four treatment groups.

Mortality was high in the very high dose groups. Among the females, only 19 animals were still alive after 78 weeks, and all surviving animals were killed at week 95. Among the males, 35 animals were still alive after 78 weeks, and 24 were still alive at 90 weeks.

During the administration period all animals were checked daily until the end of the treatment period. Additionally, once each month (and once every two weeks in the second year of the study) all animals received a detailed clinical examination for viability and clinical observations. A complete histopathological examination was performed on all animals from all groups found dead, killed moribund, or sacrificed during or at the end of the experiment.

2.1 Sponsor's analyses

2.1.1 Survival analysis

The sponsor conducted Armitage-Cochran tests of trend [1] for survival times across dose groups. In males, the trend was found to be significant only when the very high dose group was included. In females, the trend test was strongly significant ($p < 0.0001$) when all four treated groups were considered, and still significant ($p = 0.0363$ or 0.0048, depending on the scores used) when the very high dose group was excluded from the analysis.

2.1.2 Tumor data analysis

The sponsor conducted an Armitage-Cochran [1] test of trend and Peto test [8] of incidence rates for each individual type of tumor.

For the male animals, two separate sets of calculations were conducted, comparing the treated groups with each of the two control groups individually. A number of statistically significant results were found (Table 4–9 in the submission), but were based on counts of no more than one animal per dose group[1]. The sponsor comments that as there is no tumor type and dose group for which the number of lesions is above 5%, no biological significance should be inferred from any of these results.

For the female animals, animals in the very high dose group were compared with the second control group; the other animals were tested against the first control. It is not clear from the sponsor's documentation which groups are used to calculate trend statistics. The only signal that the sponsor finds noteworthy is the increase in the incidence of thyroidal c-cell adenomas, although they observe that the p-value of 0.018 is not significant at the $\alpha = 0.01$ significance level. The sponsor also notes a small p-value for uterine epithelial hyperplasias[2] ($p = 0.051$ when the treated animals are compared with the first control, and 0.068 when compared with the second control). Such tumors were developed by 10 animals in each of the control groups, and 7, 8, 16 and 12 in the low, mid, high, and very high dose groups respectively.

2.2 Reviewer's analysis

To verify sponsor's analyses and to perform additional analysis 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 all the following analyses, the two identical control groups were combined.

2.2.1 Survival analysis

Survival distributions for all five groups of animals were estimated by means of the Kaplan-Meier product limit method. The dose response relationship and the homogeneity of survival distributions were tested using the log-rank test. The intercurrent mortality data are given in Tables 1 and 2 for male and female rats, respectively. The Kaplan-Meier curves for the survival rates are given in Figures 1 and 2 for male and female rats, respectively. The results of the log-rank tests (for trend and heterogeneity) the of intercurrent mortality data are presented in Tables 3 and 4.

Reviewer's findings This reviewer's analysis found strong evidence of a trend between survival times and doses, both in the male rats and the female rats. This evidence was weakened, but remained statistically significant when the very high dose group was removed from the analysis. For both male and female rats, there was also strong evidence of heterogeneity of survival times between the five treatment groups. However, when the very high dose groups were removed, the evidence of heterogeneity among treatment groups was only statistically significant for the female

[1] These results were not found to be statistically significant under the Poly-3 analysis conducted by the reviewer

[2] Note that such tumors are not neoplastic, and so are outside the scope of this review.

Table 1: Intercurrent mortality data for female rats

	0 mg/kg/day		0.05 mg/kg/day		0.15 mg/kg/day		0.5 mg/kg/day		2.5 mg/kg/day	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	5	4.90	1	1.96	4	7.84	2	3.92	17	33.33
53 - 78	4	8.82	5	11.76	3	13.73	7	17.65	15	62.75
79 - 91	10	18.63	5	21.57	4	21.57	7	31.37	3	68.63
92 - 104	49	66.67	8	37.25	4	29.41	10	50.98	16	100.00
Ter. Sac.	34	33.33	32	62.75	36	70.59	25	49.02	.	.

Table 2: Intercurrent mortality data for male rats

	0 mg/kg/day		0.05 mg/kg/day		0.15 mg/kg/day		0.5 mg/kg/day		2.5 mg/kg/day	
Week	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %	No. of Death	Cum. %
0 - 52	2	1.96	.	.	1	1.96	3	5.88	6	11.76
53 - 78	5	6.86	3	5.88	4	9.80	2	9.80	9	29.41
79 - 91	8	14.71	3	11.76	1	11.76	1	11.76	10	49.02
92 - 104	5	19.61	4	19.61	3	17.65	7	25.49	8	64.71
Ter. Sac.	82	80.39	41	80.39	42	82.35	38	74.51	18	35.29

Table 3: Hypothesis tests for female rat intercurrent mortality data

	χ^2 statistic	Degrees of freedom	p-value
Test of Trend	70.147	4	< 0.0001
Heterogeneity	72.161	1	< 0.0001

Table 4: Hypothesis tests for male rat intercurrent mortality data

	χ^2 statistic	Degrees of freedom	p-value
Test of Trend	52.497	4	< 0.0001
Heterogeneity	53.186	1	< 0.0001

Figure 1: Survival plots for female rats

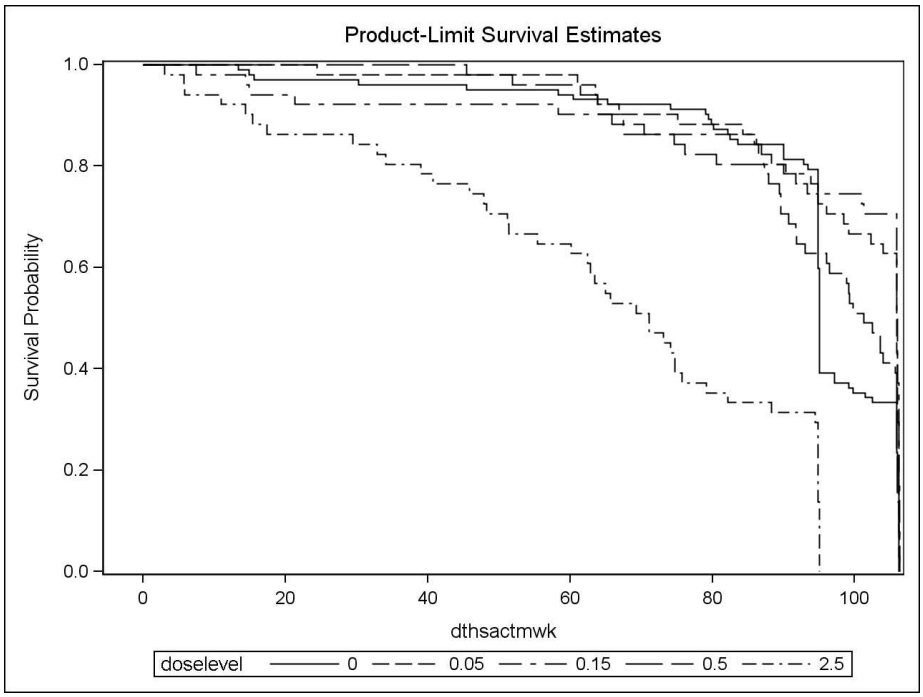
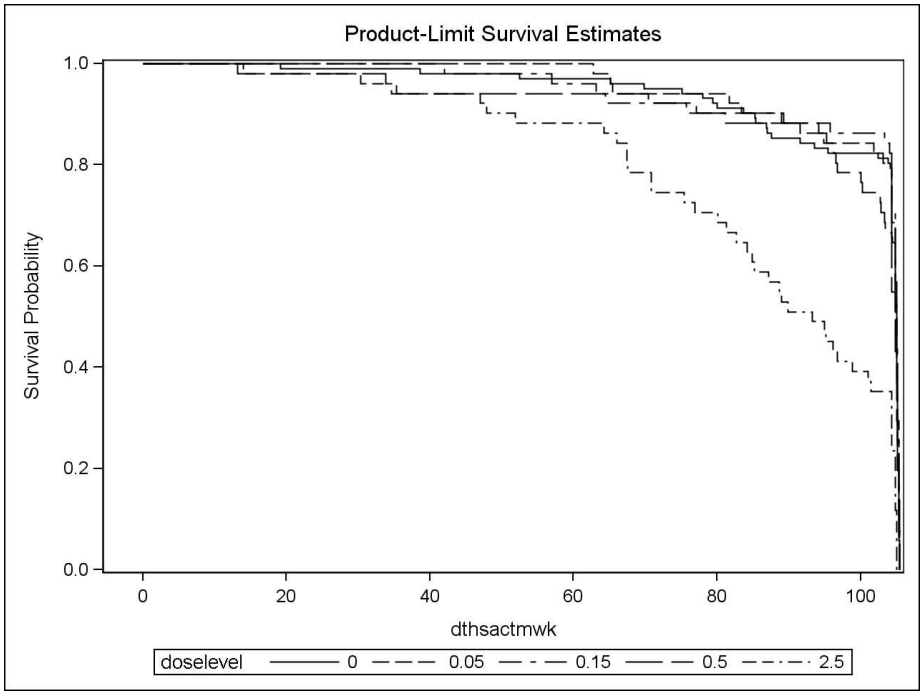


Figure 2: Survival plots for male rats



rats. When each treatment group was compared directly against the combined control, a significant difference in survival times was found for the high and very high dosage groups (for the female rats), but only for the very high dosage group (for the male rats).

For neither male nor female rats were there statistically significant differences in survival times between animals in the two control groups.

2.2.2 Tumor data analysis

Theoretical underpinnings The tumor data were analyzed for dose response relationships and pairwise comparisons of tumor incidence of combined control 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[2] and developed in the paper of Bieler and Williams[3]. In this method an animal h that lives the full study period (w_m) or dies before the terminal sacrifice with at least one tumor 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

$$s_h = \left(\frac{w_h}{w_m}\right)^k < 1.$$

The adjusted group size is defined as $\sum_h s_h$. 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 $\sum s_h$ is equal to N (the original group size) if all animals live up to the end of the study or if each animal 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 relationship between tumor onset time and 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 5 and 6 for male and female rats, respectively.

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 significance levels $\alpha = 0.005$ for common tumors and $\alpha = 0.025$ for rare tumors for a submission with two species, and a significance level $\alpha = 0.01$ for common tumors an $\alpha = 0.05$ for rare tumors for a submission with one species study 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 $\alpha = 0.01$ for common tumors and $\alpha = 0.05$ 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 [7]. In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Rahman and Lin [9] showed that this rule for multiple testing for dose response relationship is also suitable for poly-k tests.

Reviewer's findings All individual tumor types reported in the data were tested. The analyses were conducted with all four dose groups, and the combined control. The tumor types that showed p-values less than or equal to 0.05 either for dose response relationship and/or pairwise comparisons of control and treated groups for at least one sex are shown in Table 7.

In the case of the female rats, for no individual organ/tumor combination was there a significant difference in incidence rate between any of the treatment groups and the control. Likewise, for no individual organ/tumor combination was there a significant trend between the incidence rate and the dose level.

Table 5: Tumor incidence for female rats

IND 22527
Dose Response Relationship Test and Pairwise Comparisons
Using Poly-3 test
Female rats (using combined control)

Organ Name	Tumor Name	0 mg Cont N=102	0.05 mg Low N=51	0.15 mg Med N=51	0.5 mg High N=51	2.5 mg V. High N=51	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H	P_Value C vs. V.H.
ADRENAL GL	ADENOMA, CORTICAL	2	1	2	3	0	0.5458	0.7358	0.4374	0.2055	1.0000
	TUMOR, MEDULLARY, BE	1	0	1	0	0	0.6986	1.0000	0.5794	1.0000	1.0000
	TUMOR, MEDULLARY, MA	1	1	0	0	0	0.8753	0.5864	1.0000	1.0000	1.0000
BONE	OSTEOSARCOMA	0	0	0	1	0	0.2603	.	.	0.3362	.
BRAIN	MENINGIOMA, BENIGN,	0	1	0	0	0	0.6484	0.3583	.	.	.
	OLIGODENDROGLIOMA MA	0	1	0	0	0	0.6484	0.3583	.	.	.
	PINEALOMA, MALIGNANT	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
EYES	MELANOMA, MALIGNANT	0	0	0	0	1	0.0864	.	.	.	0.1979
HEART	METASTATIC SARCOMA,	0	1	0	0	0	0.6484	0.3583	.	.	.
KIDNEYS	LIPOSARCOMA	0	1	0	0	0	0.6484	0.3583	.	.	.
LACRIMAL GL	ADENOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
LIVER	ADENOMA, HEPATOCELLU	0	2	0	0	0	0.7387	0.1265	.	.	.
	CARCINOMA, HEPATOCEL	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
MAMMARY AREA	ADENOCARCINOMA	6	2	1	0	0	0.9946	0.8435	0.9567	1.0000	1.0000
	ADENOMA	3	1	1	0	0	0.9395	0.8321	0.8264	1.0000	1.0000
	FIBROADENOMA	8	2	4	3	1	0.6704	0.9253	0.6553	0.7845	0.8692
	FIBROMA, PARTIALLY N	1	1	0	1	0	0.5283	0.5864	1.0000	0.5575	1.0000
MESENT. LYMPH N	LYMPHANGIOMA	0	1	0	0	0	0.6484	0.3583	.	.	.
NASAL CAVITY	FIBROMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
OVARIES	ADENOMA, TUBULOSTROM	2	1	0	2	0	0.5007	0.7395	1.0000	0.4126	1.0000
	CYSTADENOCARCINOMA	0	1	0	0	0	0.6484	0.3583	.	.	.
	CYSTADENOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	THECOMA, BENIGN	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	TUMOR, GRANULOSA CEL	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	TUMOR, SERTOLI CELL,	0	1	0	1	0	0.3279	0.3583	.	0.3362	.
	TUMOR, SEX CORD STRO	0	1	0	0	0	0.6484	0.3583	.	.	.
PANCREAS	ADENOMA, ISLET CELL	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
PITUITARY GL	ADENOMA OF PARS DIST	41	26	18	14	5	0.9855	0.1815	0.7892	0.9484	0.9803
	ADENOMA OF PARS INTE	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	CARCINOMA OF PARS DI	0	1	0	0	0	0.6484	0.3583	.	.	.
	CARCINOMA OF PARS IN	0	0	0	1	0	0.2603	.	.	0.3362	.
SALIVARY GL	ADENOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
SKIN	CARCINOMA, SQUAMOUS	0	1	0	0	0	0.6484	0.3583	.	.	.
	FIBROMA, SUBCUTANEOU	0	0	0	1	0	0.2603	.	.	0.3362	.
	FIBROSARCOMA	0	0	0	1	0	0.2636	.	.	0.3419	.
	PAPILLOMA, SQUAMOUS	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	TUMOR, BASAL CELL, B	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
SMALL INTESTINE	LEIOMYOSARCOMA	0	1	0	0	0	0.6484	0.3583	.	.	.
STOMACH	PAPILLOMA, SQUAMOUS	0	0	0	1	0	0.2603	.	.	0.3362	.
SYSTEMIC	MALIGNANT LYMPHOMA	2	1	1	1	1	0.2524	0.7395	0.7395	0.7114	0.4880
	SARCOMA, HISTIOCYTIC	0	0	1	0	0	0.4545	.	0.3583	.	.
THYMUS	THYMOMA, BENIGN	6	6	3	4	0	0.8947	0.2263	0.6708	0.4407	1.0000
THYROID GL	ADENOMA, C-CELL	6	1	2	5	1	0.4138	0.9582	0.8400	0.2813	0.7943
	ADENOMA, FOLLICULAR	7	1	1	2	0	0.8835	0.9729	0.9729	0.8632	1.0000
	CARCINOMA, C-CELL	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	CARCINOMA, FOLLICULA	2	2	1	0	0	0.9047	0.4470	0.7358	1.0000	1.0000
TONGUE	TUMOR, GRANULAR CELL	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
UTERUS	ADENOCARCINOMA	2	2	3	4	1	0.2951	0.4470	0.2397	0.1000	0.4840
	ADENOMA	1	1	0	1	0	0.5311	0.5902	1.0000	0.5613	1.0000
	CARCINOMA, SQUAMOUS	0	0	0	0	1	0.0864	.	.	.	0.1979
	HEMANGIOPERICYTOMA,	0	1	0	0	0	0.6484	0.3583	.	.	.
	SCHWANNOMA, MALIGNAN	1	2	1	0	0	0.8209	0.2874	0.5864	1.0000	1.0000
	TUMOR, GRANULAR CELL	0	0	1	0	0	0.4521	.	0.3529	.	.
ZYMBAL'S GL	CARCINOMA, SQUAMOUS	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6: Tumor incidence for male rats

IND 22527
Dose Response Relationship Test and Pairwise Comparisons
Using Poly-3 test
Male rats (using combined control)

Organ Name	Tumor Name	0 mg Cont N=102	0.05 mg Low N=51	0.15 mg Med N=51	0.5 mg High N=51	2.5 mg V. High N=51	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H	P_Value C vs. V.H.
ADRENAL GL	ADENOMA	3	3	1	3	3	0.1087	0.3268	0.8068	0.3074	0.1975
	CARCINOMA, CORTICAL	0	0	0	0	1	0.1288	.	.	.	0.2698
	TUMOR, MEDULLARY, BE	1	0	0	2	0	0.3981	1.0000	1.0000	0.2511	1.0000
	TUMOR, MEDULLARY, MA	1	0	1	0	0	0.7216	1.0000	0.5572	1.0000	1.0000
AXILLARY LN	HEMANGIOSARCOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
BODY CAVITIES	SCHWANNOMA	1	0	0	1	0	0.5060	1.0000	1.0000	0.5507	1.0000
BRAIN	ASTROCYTOMA, MALIGNA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	GLIOMA, MIXED, BENIG	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	TUMOR, GRANULAR CELL	1	0	0	1	0	0.5060	1.0000	1.0000	0.5507	1.0000
EPIDIDYMIDES	MESOTHELIOMA, BENIGN	0	1	0	0	0	0.6502	0.3381	.	.	.
EYES	MELANOMA, MALIGNANT	0	0	0	0	1	0.1255	.	.	.	0.2640
FEMUR/MARROW	OSTEOSARCOMA	0	1	0	0	0	0.6502	0.3381	.	.	.
HEART	SCHWANNOMA, ENDOCARD	0	0	1	0	0	0.4715	.	0.3333	.	.
KIDNEYS	RENAL LIPOSARCOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
LIVER	ADENOMA, HEPATOCELLU	2	0	1	1	0	0.6601	1.0000	0.7070	0.7004	1.0000
	CARCINOMA, HEPATOCEL	0	0	0	1	0	0.2966	.	.	0.3285	.
LUNGS	ADENOMA, BRONCHIO-AL	2	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	METASTATIC OSTEOSARC	0	0	1	0	0	0.4715	.	0.3333	.	.
MAMMARY AREA	FIBROADENOMA	0	0	1	0	0	0.4715	.	0.3333	.	.
MEDIASTINUM	OSTEOSARCOMA	0	0	1	0	0	0.4715	.	0.3333	.	.
MESENT. LN	HEMANGIOMA	2	0	0	2	1	0.1841	1.0000	1.0000	0.3987	0.6048
PANCREAS	ADENOMA, ISLET CELL	5	3	1	4	0	0.8709	0.5469	0.9170	0.3343	1.0000
	ISLET CELL CARCINOMA	0	1	0	0	0	0.6502	0.3381	.	.	.
PARATHYROID GL	ADENOMA	5	2	3	2	1	0.6865	0.7522	0.5348	0.7344	0.8480
	CARCINOMA	0	0	0	1	0	0.2966	.	.	0.3285	.
PITUITARY GL	ADENOCARCINOMA, PARS	1	1	0	0	0	0.8785	0.5635	1.0000	1.0000	1.0000
	ADENOMA OF PARS DIST	18	11	11	9	7	0.5106	0.3865	0.3865	0.5723	0.5290
	ADENOMA OF PARS INTE	2	0	1	3	0	0.5814	1.0000	0.7133	0.1990	1.0000
PREPUTIAL GL	CARCINOMA, SQUAMOUS	0	1	0	1	1	0.0993	0.3381	.	0.3285	0.2640
PROSTATE	ADENOMA	2	3	0	3	1	0.3814	0.2134	1.0000	0.1990	0.6048
SALIVARY GL	ADENOCARCINOMA	0	0	0	1	0	0.2966	.	.	0.3285	.
	HISTIOCYTOMA, FIBROU	0	0	1	0	0	0.4715	.	0.3333	.	.
SEMINAL VES	ADENOMA	1	0	1	0	0	0.7216	1.0000	0.5572	1.0000	1.0000
SKIN	ADENOMA, SEBACEOUS C	0	0	0	0	1	0.1255	.	.	.	0.2640
	CARCINOMA	0	1	0	0	0	0.6502	0.3381	.	.	.
	CARCINOMA, SQUAMOUS	2	0	0	0	1	0.3312	1.0000	1.0000	1.0000	0.6013
	FIBROSARCOMA	0	0	1	0	1	0.1027	.	0.3381	.	0.2640
	HISTIOCYTOMA, FIBROU	1	0	1	0	1	0.2323	1.0000	0.5572	1.0000	0.4684
	KERATOACANTHOMA	2	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	LIPOMA	0	1	0	0	0	0.6502	0.3381	.	.	.
	PAPILLOMA, SQUAMOUS	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	SARCOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
	TUMOR, BASAL CELL, B	0	1	0	0	0	0.6502	0.3381	.	.	.
	TUMOR, BASAL CELL, M	0	0	0	1	0	0.2966	.	.	0.3285	.
	TUMOR, HAIR FOLLICLE	0	1	0	0	0	0.6502	0.3381	.	.	.
SMALL INTESTINE	HISTIOCYTOMA, FIBROU	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
SPINAL CORD	ASTROCYTOMA, MALIGNA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
STOMACH	FIBROMA	0	0	1	0	0	0.4715	.	0.3333	.	.
	LEIOMYOSARCOMA	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
SYSTEMIC	MALIGNANT LYMPHOMA	2	2	1	4	2	0.1546	0.4201	0.7101	0.0977	0.2805
	SARCOMA, HISTIOCYTIC	1	0	0	1	0	0.5060	1.0000	1.0000	0.5507	1.0000
TESTES	ADENOMA, INTERSTITIA	1	1	1	0	0	0.7896	0.5635	0.5572	1.0000	1.0000
	MESOTHELIOMA, BENIGN	0	1	0	0	0	0.6502	0.3381	.	.	.
	MESOTHELIOMA, MALIGN	0	1	0	0	0	0.6502	0.3381	.	.	.
	TUMOR SERTOLI CELL M	1	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000
THYMUS	THYMOMA, BENIGN	2	0	1	3	0	0.5829	1.0000	0.7070	0.1990	1.0000
THYROID GL	ADENOMA, C-CELL	3	11	6	2	1	0.9269	<0.001*	0.0372	0.5318	0.7117
	ADENOMA, FOLLICULAR	16	5	4	3	2	0.9244	0.9065	0.9529	0.9809	0.9781
	CARCINOMA, FOLLICULA	2	1	0	0	0	0.9581	0.7133	1.0000	1.0000	1.0000
ZYMBAL'S GL	ADENOMA, SEBACEOUS C	0	0	1	0	0	0.4715	.	0.3333	.	.
	CARCINOMA, SEBACEOUS	1	0	1	1	0	0.5096	1.0000	0.5572	0.5507	1.0000
	CARCINOMA, SQUAMOUS	1	1	0	0	0	0.8785	0.5635	1.0000	1.0000	1.0000

Table 7: Tumor types with $p \leq 0.05$ for dose response of pairwise comparison in rat study

Analysis of thyroidal c-cell adenomas
NDA 22527

Study	Statistic	Trend test	Control	Low Dose	Mid Dose	High Dose	Very High Dose
Female rats	p value for test of trend	0.3984	.	.	.	.	.
	p value for comparison with control	.	.	0.9544	0.8444	0.2737	0.7824
	Number of animals with tumor	.	6	1	2	5	1
	Adjusted number of animals at risk	.	77.0904	41.8847	42.3917	38.4899	18.26684
	Tumors per 100 (adjusted) animals	.	7.783	2.388	4.718	12.99	5.4744
Male rats	p value for test of trend	0.9199	.	.	.	.	.
	p value for comparison with control	.	.	0.000400	0.03720	0.5218	0.7022
	Number of animals with tumor	.	3	11	6	2	1
	Adjusted number of animals at risk	.	92.0482	47.4133	46.4142	44.6933	32.98769
	Tumors per 100 (adjusted) animals	.	3.259	23.20	12.93	4.475	3.0314

In the case of the male rats, there was a significant difference noted in the incidence rates of c-cell adenomas of the thyroid gland: Only three of the control animals (out of a poly-3 adjusted population size of 92 animals) developed such tumors, compared with 11 (out of 47.4) in the low dose group and 6 (out of 46.4) in the mid dose group. However, the difference in incidence rates between the high and very high dose groups, compared with the control group were not statistically significant. Likewise, the test for trend did not yield a significant result. Note also that there was no significant signal for thyroidal c-cell adenomas among the female rats.

At the request of the pharmacological reviewer, additional analyses were conducted of the following composite outcomes: All uterine tumors, all malignant uterine tumors, smooth muscle uterine tumors, connective issue uterine tumors, systemic tumors, c-cell thyroidal tumors, all adenomas (irrespective of location), and all hemangiomas and hemangiosarcomas (irrespective of location). (For obvious reasons, the uterine endpoints were only analysed in the female rats. The other composite endpoints were analysed in both male and female rats.)

In the case of the smooth muscle and connective tissue uterine tumors, there were no cases in any of the rat treatment groups, or the control groups, so no inference about the relative risks of developing such tumors may be drawn. The test for thyroidal c-cell tumors yielded a significant result, among the male rats, but as there were no thyroidal C-cell tumors among the male rats that were not adenomas, this composite endpoint reduces to the single endpoint described above. (Among the female rats, only one animal (from the control group) developed a thyroidal c-cell tumor that was not an aednoma, and so the act of broadening the endpoint only serves to enlarge the p-values.)

Based on the criteria of adjustment for multiple testing discussed above, the incidence of thyroidal c-cell adenomas in the low dose group is the only noteworthy carcinogenicity signal in the rat study. Note that the absence of a significant signal in the high and very high dose groups is not evidence that no such signal exists at those higher dose levels; the numbers of cases (3 out of an adjusted population size of 92 in the control group, 2 out of an adjusted population of 44.7 in the mid dose group, and 1 out of an adjusted population size of 33 in the very high dose group) are too small to permit meaningful inference.

2.2.3 Analysis of missing organs

Both the studies in this submission were characterized by a large number of organs which were left either unexamined, or which were incorrectly recorded as such in the supplied datasets. Taken at face value, these data suggest that certain organs were almost uniformly unexamined. This information is summarized in Tables 8 and 9 and Figures 3 and 4. In such cases, the available sample sizes are so tiny as to preclude any reasonable analysis.

Among female rats, the main widely missing organ is Bone; only one female mouse was recorded as having had its bone marrow examined; that animal was found to have bone tumors.

Figure 3: Organs reported missing in at least 10 animals in at least one study

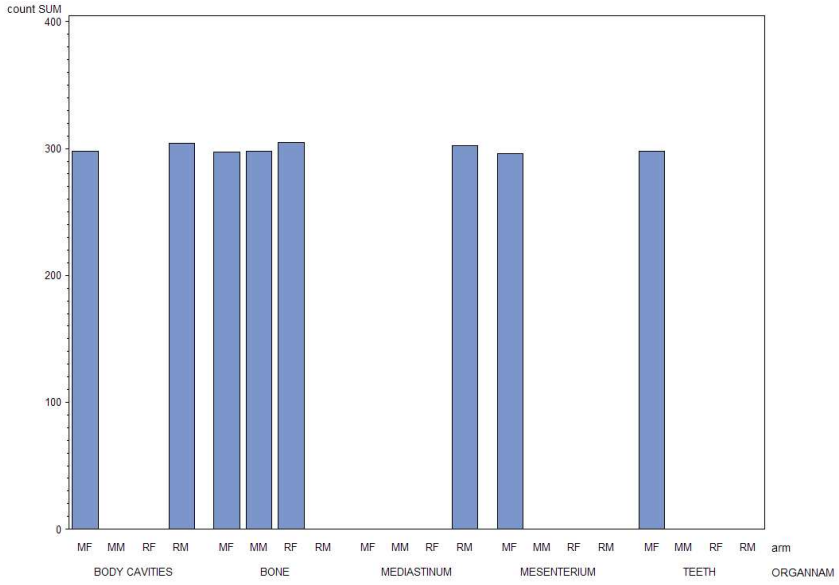


Figure 4: Organs found autolytic in at least 10 animals in at least one study

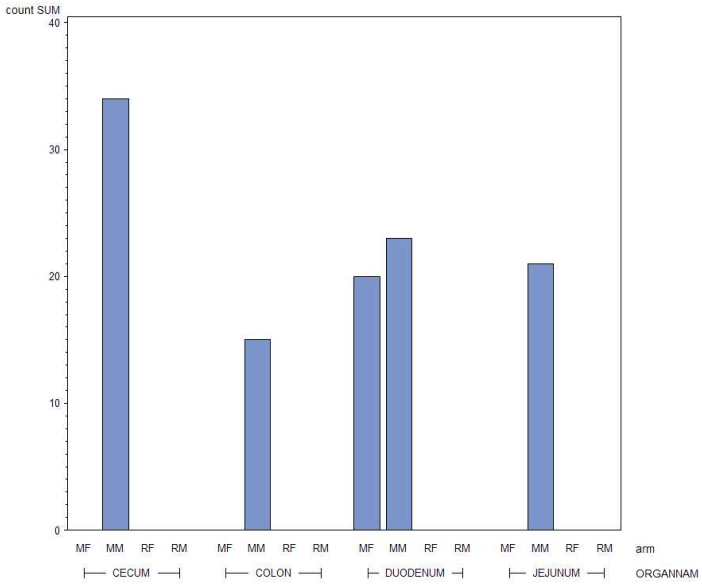


Table 8: Table of organs reported missing by arm

ORGANNAM	MF	MM	RF	RM
ADRENAL GL	1	1	0	0
AXILLARY LN	0	0	0	26
BODY CAVITIES	298	0	0	304
BONE	297	298	305	0
BRAIN	0	1	0	0
CECUM	0	6	0	0
COLON	0	10	0	0
DUODENUM	3	5	0	0
EYES	0	0	5	0
FEMUR/MARROW	5	0	0	0
HARDERIAN GL	2	2	0	0
HEART	0	0	1	0
JEJUNUM	0	2	0	0
KIDNEYS	0	0	0	0
LACRIMAL GL	0	0	3	0
MAMMARY AREA	15	4	1	1
MEDIASTINUM	0	0	0	302
MESENT. LN	26	30	0	1
MESENT. LYMPH N	0	0	5	0
MESENTERIUM	296	0	0	0
NASAL CAVITY	0	0	2	0
OVARIES	4	0	0	0
PANCREAS	7	8	2	2
PARATHYROID GL	0	0	0	19
PITUITARY GL	12	0	9	22
PREPUTIAL GL	0	21	0	20
PROSTATE	0	22	0	6
SALIVARY GL	0	0	4	0
SEMINAL VES	0	0	0	2
SKELETAL MUSCLE	1	0	0	0
SKIN	4	6	0	0
SPINAL CORD	0	1	0	0
SPLEEN	0	1	0	0
STERNUM/MARROW	3	0	0	0
STOMACH	1	0	0	0
TEETH	298	0	0	0
THYMUS	0	0	15	22
THYROID GL	0	3	4	1
TONGUE	0	0	3	0
URINARY BLADDER	0	4	0	0
UTERUS	4	0	0	0
VAGINA	26	0	0	0
ZYMBAL'S GL	0	0	33	37

Table 9: Table of autolytic organs by study

ORGANNAM	MF	MM	RF	RM
ADRENAL GL	1	0	0	0
CECUM	0	34	0	0
COLON	0	15	0	0
DUODENUM	20	23	0	0
EYES	0	0	1	0
HARDERIAN GL	2	1	0	0
JEJUNUM	0	21	0	0
MAMMARY AREA	2	0	0	0
MESENT. LN	3	1	0	0
MESENT. LYMPH N	0	0	1	0
OVARIES	2	0	0	0
PANCREAS	1	0	0	0
SKIN	1	0	0	0
SMALL INTESTINE	0	0	3	0
STOMACH	3	3	1	0
THYMUS	0	0	1	0
THYROID GL	0	0	2	0
URINARY BLADDER	0	2	0	0
UTERUS	1	0	0	0
VAGINA	1	0	0	0
ZYMBAL'S GL	0	0	0	0

In the case of male rats, there are two organs which are consistently listed as being unexamined: Body Cavities and Medisatinum. Only two male rats were listed as having their body cavities examined (one in the control group, one in the high dose group). Both were found to have had tumors in their body cavities. Only four male rats are recorded as having their mediastinum analyzed; one of these four was found to have a tumor.

3 Mouse Study

Two separate experiments were conducted, one in male and one in female mice. In each of these two experiments there were three treated groups and two identical control groups. Three hundred (Crl:CD-1^{TM}(ICR)BR) mice of each sex were randomly allocated to treated and control groups in equal size of 60 animals. The dose levels for treated groups were 0.025, 0.25, and 2.5 mg/kg/day. In this review these dose groups were referred to as the low, medium, and high dose groups respectively. The vehicle for the test was a solution of graded millipore water, in concentrations of 0.0005%, 0.005% and 0.05% for the three treatment groups.

Due to a high mortality rate, the high dose male group stopped receiving treatment at week 92.

During the administration period animals were checked at least once daily for viability and clinical observations. All animals were checked once immediately after dosing, until the end of the treatment period. Additionally, once each month (and every two weeks after week 82) all animals received a palpable mass examination. A complete histopathological examination was performed on all animals from all groups found dead, killed moribund, or sacrificed during or at the end of the experiment.

3.1 Sponsor's analyses

3.1.1 Survival analysis

Mortality data were analyzed by life-table techniques of Kaplan-Meier [6].

The survival distributions and quantiles for each group were estimated, and the two-sided Mantel-Cox log-rank test for equality and test for progressive trend were also conducted.

Among males, the two control groups were found to have comparable survival times, and so were pooled for this analysis. The high dose groups was found to experience significantly higher mortality ($p = 0.0001$) than the control group. The trend was also significant. However, the wide range of dose levels strongly affected the p-value of the slope. When the trend was calculated relative to the actual dosage, the p-value was 0.0001, but when the slope was calculated relative to the ordinal value of the groups (0 for controls, 1 for low etc.), the p-value 0.0006.

In the case of the female mice, it was observed that the two control groups had significantly different survival times ($p = 0.0405$), and so subsequent survival analyses were carried out using the better performing control group. Strongly significant differences are noted between both the mid and high dose groups ($p = 0.0044$ and $p = 0.0010$ respectively) and the controls. The trend is significant when ordinal values are used, but not when absolute dosages are used ($p = 0.0605$). The sponsor considers this to be evidence of a nonlinear dose response.

3.1.2 Tumor data analysis

For each tumor type, either a trend test (using the Armitage-Cochran [1] method) was conducted (in the case of non-neoplastic lesions), or an analysis using Peto's method [8] was conducted (in the case of neoplastic lesions). For each tumor type two sets of statistics were found, one based on contrast with the first control group, and one based on contrast with the second control group.

Male mice: Statistically significant values for malignant lymphoma, histiocytic sarcoma, hepatic hemangiosarcoma, duodenal adenoma and colonic adenocarcinoma were obtained.

The most striking result was for malignant lymphoma, where significant p-values (less than 0.0005) were found when contrasting both the mid and high dose groups with the controls. The sponsor suggests that this result was to be expected as a result of the pharmacological action of FTY720. The sponsor also claims that there is no difference in the distribution of different types of malignant lymphomas between the dose groups and the controls (see Table 4-8 in the sponsor's submission), although the numbers of cases are sufficiently small that such an analysis would be crude at best. The sponsor also suggests that "the high incidence at 2.5 mg/kg could be related to an overdosage consistent with the high mortality rate".

Statistically significant values were obtained for the occurrence of hemangiosarcoma in the liver ($p = 0.008$ and $p = 0.0008$) in the mid and high dose groups (relative to the control). Furthermore, when all hemangiosarcomas were pooled, a significant difference was detected between the high dose group and the controls. However, the sponsor argues that the overall incidence rate in the high dose group (6 animals out of 60) is within the normal range given by the historical control data for this species of mouse.

Statistically significant values were also obtained for the occurrence of histiocytic sarcoma ($p = 0.0205$ for mid dose vs control and $p = 0.0066$ for high dose vs control). However, the sponsor considers any observed incidence rate less than 5% to be inadequate to provide eivdence of a biologically relevant effect.

The other statistically significant tumor types were instances where no more than one tumor in each dose group was detected.[3]

Female Mice: As with the male mice, the sponsor reported a treatment-related increase in malignant lymphoma at the doses of 0.25 and 2.5 mg/kg. However, this increase was not significant

[3] As in the case of the female rats, these cases were not found to be significant according to a poly-3 analysis.

at the $\alpha = 0.01$ level deemed appropriate for common neoplasms. Immunohistochemical stainings characterizing B and T-cell lymphocytes revealed no obvious difference in the type of malignant lymphoma between the spontaneous lymphomas of the control groups and that of the high dose group (see Table 4-10 of the sponsor's submission).

Statistically significant values were noted for the occurrence of adenocarcinoma in the harderian gland, adenoacanthoma in the mammary gland, fibroma in the uterus, benign medullary tumor in the adrenals, and hemangioma in the pancreas. However, for each of these cases, no more than one animal per treatment group was found to have developed a tumor, and so the sponsor concluded that the results were of no biological relevance.

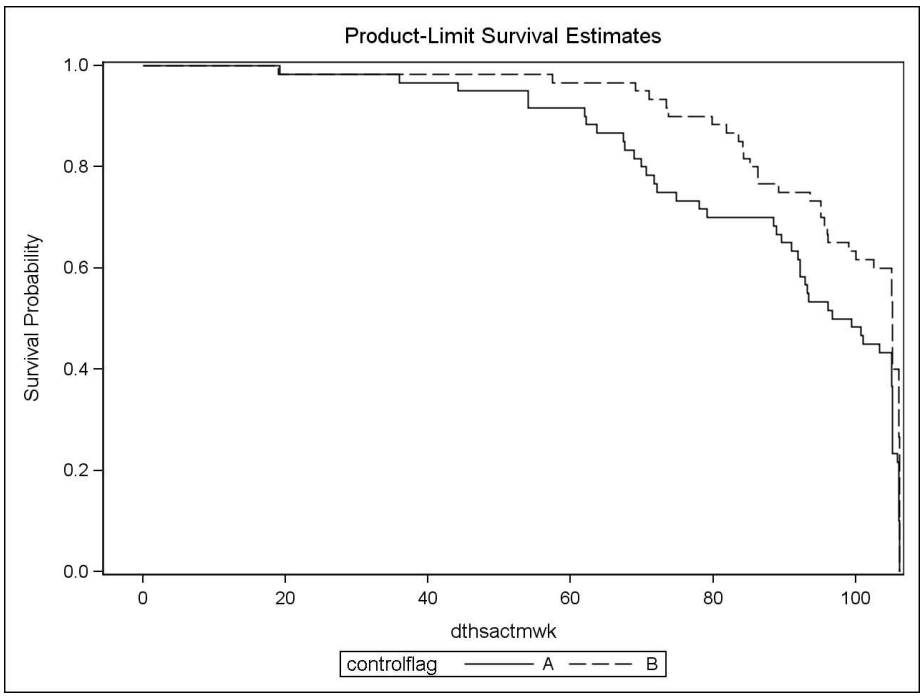
Pooling hemangiomas over the whole body revealed a statistically significant value with an increased number of animals (4/60) presenting hemangiomas at 2.5 mg/kg. However, no distinct statistically significant value was obtained by pooling hemangiomas plus hemangiosarcomas over the whole body and there was no trend by pooling hemangiosarcomas alone or angiomatous hyperplasias over the whole body. The fact that these trends were not significant, taken together with claimed background rates for this species of mice, led the sponsor to conclude that these results were not biologically relevant.

3.2 Reviewer's analysis

This reviewer independently performed survival and tumor data analyses from the mouse study. For the mouse data analyses this reviewer used similar methodologies as he used to analyze the data from the rat study. Data used in this reviewer's analyses were provided by the sponsor electronically.

A significant difference was found between the two control groups in study of female mice, with group 0 experiencing a higher mortality rate than group 4. The Kaplan-Meier plots for the control groups are shown as Figure 5. Absent any explanation for this discrepancy, the conservative approach of basing all subsequent survival analyses (although not the tumor analyses) on the better performing (second) control group has been adopted.

Figure 5: Survival plots for female mice – control groups



When testing two control groups, it is reasonable to assume proportional hazards under the null hypothesis, so rather than a log-rank test, a maximum likelihood proportional hazards test has been used in this case. The estimate of the hazard ratio between the two control groups is 1.337, with a confidence interval of $(0.145, 1.562)$. The p value is 0.0002. Such a strong divergence between the two control groups is striking and raises concerns that the study has been insufficiently controlled.

(When a log-rank test is conducted, the difference between the two control groups is less pronounced, but still statistically significant ($p = 0.0312$). This weakening of effect is not surprising, as the log-rank test is a non parametric test (compared to the maximum likelihood test, which is semiparametric), and so generally more conservative. However, the only justification for preferring the log-rank test here would be if it was felt that the assumption of proportional hazards was not valid, in which case one must already have grave doubts about the interchangeability of the two control groups.)

Analyses of the male mice were conducted using the combined control.

3.2.1 Survival analysis

Survival distributions for all four groups of animals were estimated by means of the Kaplan-Meier product limit method. The dose response relationship and the homogeneity of survival distributions were tested using the log-rank test. The intercurrent mortality data are given in Tables 10 and 11 for male and female mice, respectively. The Kaplan-Meier curves for the survival rates are given in Figures 6 and 7 for male and female mice, respectively. The results of log-rank tests of intercurrent mortality data are presented in Tables 12 and 13.

Table 10: Intercurrent mortality data for female mice

	0 mg/kg/day		0.025 mg/kg/day		0.25 mg/kg/day		2.5 mg/kg/day	
	No. of		No. of		No. of		No. of	
Week	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %
0 - 52	1	1.67	1	1.67	6	10.00	3	5.00
53 - 78	5	10.00	10	18.33	13	31.67	15	30.00
79 - 91	9	25.00	12	38.33	11	50.00	11	48.33
92 - 104	9	40.00	8	51.67	10	66.67	7	60.00
Ter. Sac.	36	60.00	29	48.33	20	33.33	24	40.00

Reviewer's findings A log-rank analysis was conducted of time to death.

This reviewer's analysis showed statistically significant dose response relationships in mortality across treatment groups in both sexes, together with statistically significant heterogeneity between the different dose groups for both sexes. The pairwise comparisons in male mice showed statistically significant increased mortality in the high dose group compared to the combined control. Similar pairwise comparisons in female mice showed statistically significant increased mortality in the mid and high dose groups compared to control group 4 (the better performing control group).

3.2.2 Tumor data analysis

The tumor rates and the p-values of the tumor types tested for dose response relationship and pairwise comparisons of control and treated groups are given in Tables 14 and 15 for male and female mice, respectively. The same composite endpoints (all uterine tumors, all malignant uterine

Table 11: Intercurrent mortality data for male mice

	0 mg/kg/day		0.025 mg/kg/day		0.25 mg/kg/day		2.5 mg/kg/day	
	No. of		No. of		No. of		No. of	
Week	Death	Cum. %	Death	Cum. %	Death	Cum. %	Death	Cum. %
0 - 52	6	5.00	4	6.67	2	3.33	6	10.00
53 - 78	25	25.83	17	35.00	18	33.33	26	53.33
79 - 91	13	36.67	4	41.67	4	40.00	7	65.00
92 - 104	35	65.83	10	58.33	20	73.33	5	73.33
Ter. Sac.	41	34.17	25	41.67	16	26.67	16	26.67

Table 12: Hypothesis tests for female mouse intercurrent mortality data

	χ^2 statistic	Degrees of freedom	p-value
Test of Trend	3.266	3	0.0362
Heterogeneity	13.167	1	0.0043

Table 13: Hypothesis tests for male mouse intercurrent mortality data

	χ^2 statistic	Degrees of freedom	p-value
Test of Trend	15.773	3	< 0.0001
Heterogeneity	16.245	1	0.0010

Figure 6: Survival plots for female mice

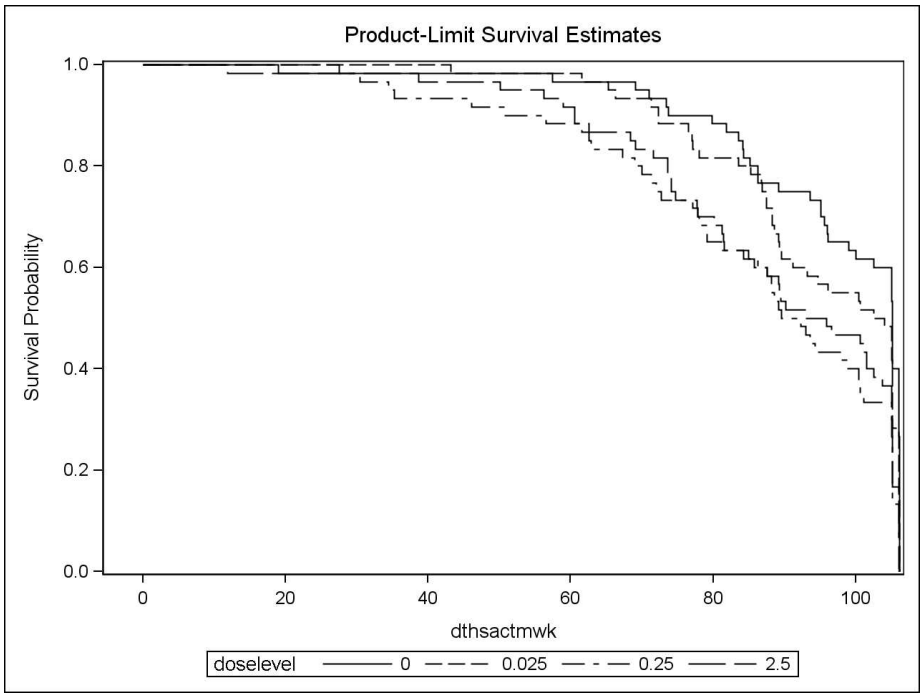
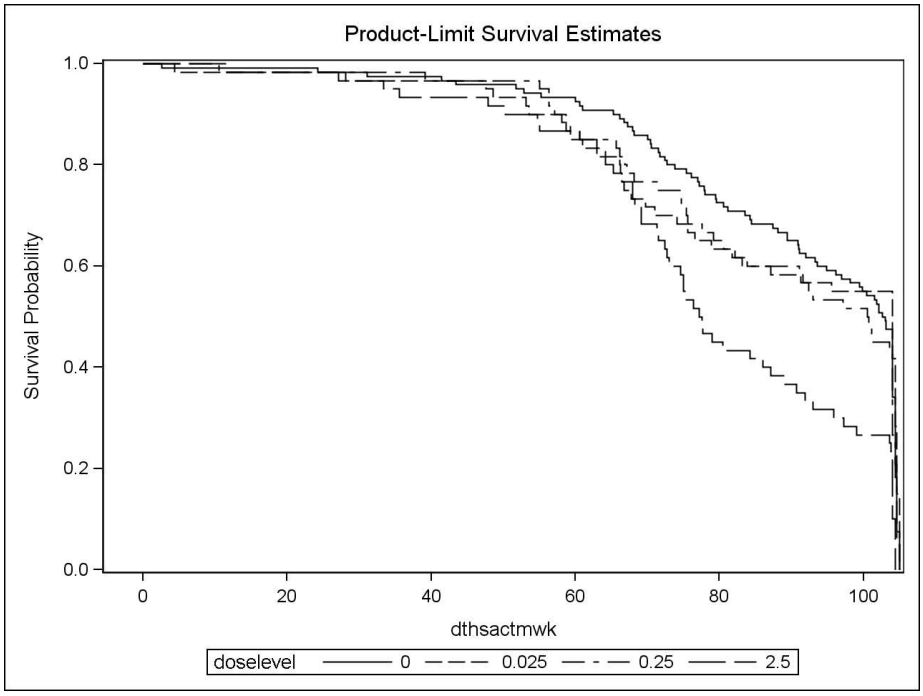


Figure 7: Survival plots for male mice



tumors, smooth muscle uterine tumors, connective issue uterine tumors, systemic tumors, c-cell thyroidal tumors, all adenomas (irrespective of location), and all hemangiomas and hemangiosarcomas (irrespective of location)) as were investigated in the rat study were also tested.

Reviewer's findings Table 16 details the tumor types and composite endpoints for which a p-value less than or equal to 0.05 was found for either the dose response or the test of proportion for at least one treatment group vs control, in at least one sex.

Based on the same multiple testing adjustment procedure discussed in the Section 2.2.2, the incidence of hepatic hemangiosarcomas in male mice shows a significant dose response trend, and a significant difference in incidence between the high dose group and the controls ($p = 0.0051$, where the threshold for a common tumor is $\alpha = 0.05$). However, no corresponding signal is detectable among the female mice.

The incidence of malignant lymphoma in male mice also displays a strongly significant trend across dose groups. Viewed in this context, the trend in the female mice ($p = 0.0293$) can be seen as providing strong corroboratory evidence for this dose response. Likewise, the difference in incidence between the control animals and the high dose animals is strongly significant in the male mice, and while the differences in the incidence rates between the male mid dose group and the control group, and between the female mid and high dose groups and the control group all fail to meet the required $p = 0.005$, they are all significant at the $\alpha = 0.05$ level, again providing strong corroboratory evidence for a carcinogenic effect in mice.

While significant when viewed individually, the difference in incidence of myeloid leukemia between the male mid dose group and the male control group fails to be significant after the multiplicity adjustment.

The interpretation of the signal for systemic tumors is similar to the interpretation (above) of the signal for malignant lymphomas, a fact which is hardly surprising since malignant lymphomas comprise over 80% of the systemic tumors reported.

3.3 Analysis of missing and autolytic organs

As with the rat study, many mice have had organs recorded as being unexamined. In the mouse study, a high incidence of autolytic organs has also been reported. The details are reported in Tables 9 and 8, and Figures 4 and 3 , but it is worth noting here that the high reported rates (98% or higher) of certain unexamined organs mean that meaningful analysis of carcinigenicity is not possible in the following cases:

> For female mice: Body cavities, Bone, Mesenterium, Teeth
> For male mice: Bone.

Among both male and female mice, there was a high incidence of autolysis. Noticably this was concentrated in a few organs. Among the female mice, the only organ which was found to be severely autolyzed (to the extent that a histopathological analysis was not possible) in more than 5% of the animals was the Duodenum (autolyzed in 20 animals). Among the male mice, the Cecum (34 animals), Duodenum (23 animals), Jejunum (21 animals) and Colon (15 animals) were all reported as being autolytic in at least 5% of animals.

4 Evaluation of the validity of the study

4.1 Suitability of dose levels

4.1.1 Issues of concern when selecting dose levels

The decision of an appropriate dose level for the high dose group is made difficult by the need to satisfy two competing imperatives: on the one hand, if the dose level is insufficiently high, then genuine carcinogenicity effects may not be apparent, but on the other hand, if the dose level is too

Table 14: Tumor incidence for female mice

IND 22527
Dose Response Relationship Test and Pairwise Comparisons
Using Poly-3 test
Female mice (using second control)

Organ Name	Tumor Name	0 mg Cont N=120	0.025 mg Low N=60	0.25 mg Med N=60	2.5 mg High N=60	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H
ADRENAL GL	ADENOMA, CORTICAL, S	2	1	0	1	0.6347	0.8671	1.0000	0.8387
	TUMOR, MEDULLARY, BE	1	0	0	1	0.4106	1.0000	1.0000	0.7009
BODY CAVITIES	MESOTHELIOMA	0	0	1	0	0.4520	.	0.4382	.
BONE	HEMANGIOSARCOMA, VER	0	0	1	0	0.4520	.	0.4382	.
	OSTEOSARCOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
DUODENUM	ADENOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
FEMUR/MARROW	HEMANGIOMA	0	0	0	1	0.2316	.	.	0.4505
	SARCOMA, NOT OTHERWI	0	1	0	0	0.7175	0.4845	.	.
HARDERIAN GL	ADENOCARCINOMA	0	0	0	1	0.2316	.	.	0.4505
	ADENOMA	2	3	2	1	0.7299	0.4704	0.5926	0.8387
LIVER	ADENOMA, HEPATOCELLU	0	1	1	0	0.5759	0.4845	0.4382	.
	CARCINOMA, HEPATOCEL	0	0	1	0	0.4520	.	0.4382	.
	CHOLANGIOMA	0	1	0	0	0.7175	0.4845	.	.
	HEMANGIOMA	0	1	1	2	0.0926	0.4845	0.4382	0.2002
	HEMANGIOSARCOMA	0	0	0	1	0.2316	.	.	0.4505
LUNGS	ADENOMA, BRONCHIO-AL	11	11	9	10	0.3882	0.5744	0.5810	0.4950
	CARCINOMA, BRONCHIO-	10	5	6	4	0.8253	0.9414	0.7861	0.9525
MAMMARY AREA	ADENOACANTHOMA	0	0	0	1	0.2316	.	.	0.4505
	ADENOCARCINOMA	0	3	1	1	0.4846	0.1137	0.4382	0.4565
	CARCINOMA IN SITU, D	1	0	0	0	1.0000	1.0000	1.0000	1.0000
MESENT. LN	HEMANGIOMA	0	1	0	0	0.7175	0.4845	.	.
MESENTERIUM	HEMANGIOMA	0	0	0	1	0.2316	.	.	0.4505
OVARIES	ADENOMA, TUBULOSTROM	0	1	1	1	0.2506	0.4845	0.4382	0.4505
	CYSTADENOMA	2	3	2	1	0.7299	0.4704	0.5926	0.8387
	GRANULOSA CELL TUMOR	0	1	0	0	0.7175	0.4845	.	.
	LUTEOMA, BENIGN	1	0	0	1	0.4106	1.0000	1.0000	0.7009
	SERTOLI CELL TUMOR,	0	1	0	0	0.7175	0.4845	.	.
	TUMOR, SERTOLI CELL	1	0	0	0	1.0000	1.0000	1.0000	1.0000
PANCREAS	HEMANGIOMA	0	0	0	1	0.2360	.	.	0.4565
SKELETAL MUSCLE	HEMANGIOMA	0	0	0	1	0.2360	.	.	0.4565
SKIN	FIBROSARCYXOMATOUS	1	0	0	1	0.4106	1.0000	1.0000	0.7009
	HISTIOCYTOMA	0	1	1	0	0.5759	0.4845	0.4382	.
	LIPOSARCOMA	0	0	1	0	0.4520	.	0.4382	.
	SARCOMA, HISTIOCYTIC	1	0	0	1	0.4173	1.0000	1.0000	0.7074
STERNUM/MARROW	TUMOR, MAST CELL, MA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
STOMACH	ADENOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
SYSTEMIC	HISTIOCYTIC SARCOMA	2	3	1	1	0.7437	0.4610	0.8227	0.8342
	MALIGNANT LYMPHOMA	14	12	25	24	0.0186	0.7138	0.0067*	0.0185
	MYELOID LEUKEMIA, GR	2	0	1	1	0.4995	1.0000	0.8286	0.8342
TEETH	CARCINOMA, SQUAMOUS	0	1	0	0	0.7175	0.4845	.	.
THYROID GL	ADENOMA, FOLLICULAR	0	0	1	1	0.1553	.	0.4382	0.4505
UTERUS	FIBROMA	0	0	0	1	0.2316	.	.	0.4505
	HEMANGIOSARCOMA	0	1	1	0	0.5745	0.4898	0.4382	.
	LEIOMYOMA	2	1	2	0	0.8434	0.8671	0.5926	1.0000
	LEIOMYOSARCOMA	0	2	0	1	0.3938	0.2322	.	0.4565
	SARCOMA, ENDOMETRIAL	0	1	1	0	0.5759	0.4845	0.4382	.
	TUMOR, GRANULAR CELL	2	0	1	1	0.4984	1.0000	0.8227	0.8342
VAGINA	FIBROMA	0	1	0	0	0.7175	0.4845	.	.

Table 15: Tumor incidence for male mice

IND 22527
Dose Response Relationship Test and Pairwise Comparisons
Using Poly-3 test
Male mice (using combined control)

Organ Name	Tumor Name	0 mg Cont N=120	0.025 mg Low N=60	0.25 mg Med N=60	2.5 mg High N=60	P_Value Dos Resp	P_Value C vs. L	P_Value C vs. M	P_Value C vs. H
ADRENAL GL	ADENOMA, CORTICAL	4	1	2	0	0.8487	0.8523	0.6230	1.0000
	CORTICAL ADENOCARCIN	1	0	0	0	1.0000	1.0000	1.0000	1.0000
	LIPOMA	0	1	0	0	0.5616	0.3154	.	.
BONE	HEMANGIOSARCOMA	0	1	0	1	0.1562	0.3206	.	0.2705
BRAIN	ASTROCYTALIGNANT	1	0	0	0	1.0000	1.0000	1.0000	1.0000
CECUM	ADENOCARCINOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
	LEIOMYOMA	0	1	0	0	0.5616	0.3154	.	.
COLON	ADENOCARCINOMA	0	0	0	1	0.1618	.	.	0.2705
DUODENUM	ADENOMA	0	0	0	1	0.1576	.	.	0.2645
HARDERIAN GL	ADENOMA	9	4	2	2	0.7441	0.6282	0.9130	0.8399
JEJUNUM	ADENOCARCINOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
KIDNEYS	ADENOMA, RENAL TUBUL	1	0	0	0	1.0000	1.0000	1.0000	1.0000
LIVER	ADENOMA, HEPATOCELLU	11	5	2	1	0.9561	0.6044	0.9557	0.9789
	CARCINOMA, HEPATOCEL	7	1	3	2	0.5064	0.9565	0.6790	0.7568
	HEMANGIOSARCOMA	1	0	3	5	0.0010*	1.0000	0.0948	0.0059*
LUNGS	ADENOMA, BRONCHIO-AL	23	10	9	4	0.9371	0.6392	0.7672	0.9645
	CARCINOMA, BRONCHIO-	10	4	7	3	0.5945	0.7109	0.2833	0.7342
MAMMARY AREA	ADENOCARCINOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
MESENT. LN	HEMANGIOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
PANCREAS	ADENOMA, ISLET CELL	1	1	0	0	0.8090	0.5330	1.0000	1.0000
PREPUTIAL GL	CARCINOMA, SQUAMOUS	0	0	1	0	0.3627	.	0.3206	.
PROSTATE	ADENOCARCINOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
SEMINAL VES	ADENOMA	2	1	0	0	0.9173	0.6826	1.0000	1.0000
	FIBROSARCOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
SKIN	CARCINOMA SQUAMOUS C	0	0	1	0	0.3596	.	0.3154	.
	HEMANGIOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
	PAPILLOMA, SQUAMOUS	1	0	0	0	1.0000	1.0000	1.0000	1.0000
SPINAL CORD	CHONDROMA, VERTEBRAL	1	0	0	0	1.0000	1.0000	1.0000	1.0000
	MENINGIOMA MALIGNANT	0	0	1	0	0.3627	.	0.3206	.
SPLEEN	HEMANGIOSARCOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
STOMACH	ADENOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000
SYSTEMIC	FIBROUS HISTIOCYTALI	0	1	0	0	0.5637	0.3206	.	.
	HISTIOCYTIC SARCOMA	1	0	0	2	0.0720	1.0000	1.0000	0.1848
	MALIGNANT LYMPHOMA	9	5	10	18	<0.001*	0.5036	0.0383	<0.001*
	MYELOID LEUKEMIA	1	2	4	0	0.7041	0.2375	0.0376*	1.0000
TESTES	ADENOMA, INTERSTITIA	4	1	0	0	0.9848	0.8550	1.0000	1.0000
	TUMOR, SERTOLI CELL,	1	0	0	0	1.0000	1.0000	1.0000	1.0000
THYROID GL	ADENOMA, FOLLICULAR	1	0	1	0	0.5950	1.0000	0.5401	1.0000
URINARY BLADDER	LEIOMYOSARCOMA	1	0	0	0	1.0000	1.0000	1.0000	1.0000

Table 16: Tumor types with $p \leq 0.05$ for dose response of pairwise comparison in mouse study

Analysis of hepatic hemangiosarcomas
NDA 22527

Study	Statistic	Trend test	Control	Low Dose	Mid Dose	High Dose
Female mice	p value for test of trend	0.4625	.	.	.	.
	p value for comparison with control	.	.	1	1	0.6682
	Number of animals with tumor	.	2	0	0	1
	Adjusted number of animals at risk	.	91.7345	46.0982	38.4177	40.6018
	Tumors per 100 (adjusted) animals	.	2.180	0	0	2.463
Male mice	p value for test of trend	0.0009	.	.	.	.
	p value for comparison with control	.	.	1	0.09010	0.005100
	Number of animals with tumor	.	1	0	3	5
	Adjusted number of animals at risk	.	88.2434	40.7116	40.8827	32.9727
	Tumors per 100 (adjusted) animals	.	1.133	0	7.338	15.16

Analysis of malignant lymphoma
NDA 22527

Study	Statistic	Trend test	Control	Low Dose	Mid Dose	High Dose
Female mice	p value for test of trend	0.0064	.	.	.	.
	p value for comparison with control	.	.	0.5356	0.000600	0.003200
	Number of animals with tumor	.	23	12	24	23
	Adjusted number of animals at risk	.	97.2451	49.3557	45.5205	48.0311
	Adjusted incidence rate	.	23.65	24.31	52.72	7.89
Male mice	p value for test of trend	<.0001	.	.	.	.
	p value for comparison with control	.	.	0.4947	0.03570	0.000015
	Number of animals with tumor	.	9	5	10	18
	Adjusted number of animals at risk	.	90.0812	43.2947	42.6818	40.7383
	Adjusted incidence rate	.	9.991	11.55	23.43	44.18

Analysis of myeloid leukemia
NDA 22527

Study	Statistic	Trend test	Control	Low Dose	Mid Dose	High Dose
Female mice	p value for test of trend	0.3382	.	.	.	.
	p value for comparison with control	.	.	1	1	0.5191
	Number of animals with tumor	.	1	0	0	1
	Adjusted number of animals at risk	.	91.9702	46.0982	38.4177	40.8237
	Tumors per 100 (adjusted) animals	.	1.087	0	0	2.450
Male mice	p value for test of trend	0.7013	.	.	.	.
	p value for comparison with control	.	.	0.2370	0.03740	1
	Number of animals with tumor	.	1	2	4	0
	Adjusted number of animals at risk	.	88.2305	41.9894	42.2630	31.4698
	Tumors per 100 (adjusted) animals	.	1.133	4.763	9.465	0

Analysis of systemic tumors
NDA 22527

Study	Statistic	Trend test	Control	Low Dose	Mid Dose	High Dose
Female mice	p value for test of trend	0.0101	.	.	.	.
	p value for comparison with control	.	.	0.6199	0.003100	0.007900
	Number of animals with tumor	.	31	15	26	26
	Adjusted number of animals at risk	.	100.92	50.1562	46.3589	49.5907
	Tumors per 100 (adjusted) animals	.	30.72	29.91	56.08	52.43
Male mice	p value for test of trend	<.0001	.	.	.	.
	p value for comparison with control	.	.	0.3678	0.007300	0.000017
	Number of animals with tumor	.	11	7	14	20
	Adjusted number of animals at risk	.	90.7721	44.5725	44.3956	42.2851
	Tumors per 100 (adjusted) animals	.	12.12	15.70	31.53	47.30

high, then there is a risk of non-carcinogenic toxic effects killing the animals before they have a chance to demonstrate a carcinogenicity effect.

Haseman [5] suggested that a satisfactory balance between these two imperatives has been found when the following two conditions are both satisfied:

1. Were enough animals exposed, for a sustained amount of time, to the risk of late developing tumors?

2. 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 fifty animals per treatment group. The following are some rules of thumb regarding these two issues as suggested by experts in this field:

Haseman [5] 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 considered as a sufficient number and adequate exposure. In addition Chu, Cueto and Ward [4], 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 [4], the following criteria are mentioned for dose adequacy. A high dose is considered as close to MTD if any of the criteria is met:

1. 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.

2. The administered dose is also considered an MTD if dosed animals exhibit clinical signs or severe histopathologic toxic effects attributed to the chemical.

3. In addition, doses are considered adequate if the dosed animals show a slight increased mortality compared to the controls.

4.1.2 Appropriateness of high dose in rat study

In the case of the rat study, there was uncertainty about an appropriate dose level for the high dose. Consequently, the study was conducted with four dose groups, rather than three. Thus if the 2.5 mg/kg dose level is found to be too toxic, then it is conceivable that the 0.5 mg/kg dose level may be acceptable as a substitute high dose level. However, to accommodate the extra dose group, there are only 51 animals in each treatment group, as compared to the 60 animals per group in the mouse study.

It was not possible to assess the appropriateness of these dose levels in advance of the study, the dose levels having been selected on the basis of small short term dose ranging studies. However, we can now perform a post-hoc analysis to measure these dose levels against the criteria presented above.

The survival data for the high and very high dose groups are presented in Table 17.

To test whether the putative high dose groups did indeed have a "slight increased mortality compared to the controls", a Cox proportional hazard regression model was fitted to survival times

Table 17: Percentage of survival in the high and very high dose groups at the end of weeks 52, 78, and 91 (rats)

Survival percentage (Number surviving)			
	End of 52 weeks	End of 78 weeks	End of 91 weeks
Females — very high	67% (34)	37% (19)	31% (16)
Females — high	96% (49)	82% (42)	69% (35)
Females — controls	95% (97)	91% (93)	81% (83)
Males — very high	88% (45)	371% (36)	51% (26)
Males — high	94% (48)	88% (44)	88% (45)
Males — controls	98% (100)	93% (95)	84% (86)

Table 18: Proportional hazards regression of high and very high dose groups compared with control (rats)

		High dose group vs control	Very high dose group vs control
Females	χ^2_1	0.0007	48.0263
	p-value	0.9788	< 0.0001
	HR (CI)	$0.995 \quad (0.710, 1.395)$	$3.684 \quad (2.548, 5.326)$
Males	χ^2_1	0.8298	32.5766
	p-value	0.3623	< 0.0001
	HR (CI)	$1.169 \quad (0.835, 1.637)$	$2.847 \quad (1.988, 4.077)$

for the high and very high dose groups against the control group. The results are presented in Table 18.

The percent difference in mean body weight gain from the concurrent combined control in the treatment groups in given in Table 19. Note that this the proportion of weight gain is given by the formula:

$$\frac{\Delta W_C - \Delta W_T}{\Delta W_C}$$

where ΔW_T is the mean weight gain for treatment animals, and ΔW_C is the mean weight gain for control animals.

In light of these data, we return to the question of whether either 2.5 mg/kg, or 0.5 mg/kg was an appropriate choice for the high dose.

In the case of the female rats, it seems clear that 2.5 mg/kg was in excess of the MTD. Mortality was considerably higher than the control (with a confidence interval for the hazard ratio of $(2.548, 5.326)$. Furthermore, the female rats in this group experienced 22% less weight gain than the control group. Finally, only 16 female rats were still alive after 91 weeks.

However, it is not so clear that 0.5 mg/kg is a suitable dosage to serve as the high dose group in a three dose group analysis. There are no concerns that this dosage level is too high, as 69% of the animals survived until 91 weeks. However, apart from failure to gain weight, the data suggest

Table 19: Mean proportion of weight gain relative to control group (rats)

	0.05 mg/kg	0.15 mg/kg	0.5 mg/kg	2.5 mg/kg
Females	92.2%	75.4%	76.4%	78.0%
Males	94.0%	86.2%	85.3%	74.6%

that the female rats were able to tolerate this dose level rather better than one would expect for a dosage close to the MTD. It is worth noting that the weight gain in these animals, while noticably lower than that observed in the control animals, was comparable to (and in fact slightly greater than) that of the animals in the 0.15 mg/kg dose group. Thus it does not seem reasonable to accept this dose level as acceptably close to the maximum tolerated dose on the basis of diminished weight gain alone. A judgement of whether this dose level is suitable for treatment as a high dose group therefore depends on other clinical signs and histopathological toxic effects.

It is not clear why, given concerns that 2.5 mg/kg was above the maximum tolerated dose, the substitute high dose level was *one fifth* of this dosage. We are left in a situation where we may have no reasonable high dose group for the female rats.

In the case of male rats, as 51% of the animals survived to 91 weeks, we can just accept this dose level as not being too toxic. Furthermore, given the clear evidence of toxicity among the male rats in this group (the 95% confidence interval for the hazard ratio relative to the control group is $(1.988, 4.077)$), and the diminished weight gain in this group, it is reasonable to accept this dose level as a suitable high dose level for the male rats.

4.1.3 Appropriateness of high dose in mouse study

Table 20: Percentage of survival in the high dose group at the end of weeks 52, 78, and 91 (mice)

Survival percentage (Number surviving)			
	End of 52 weeks	End of 78 weeks	End of 91 weeks
Females — high	95% (57)	70% (42)	52% (31)
Females — controls	95% (57)	72% (43)	63% (38)
Males — high	90% (54)	47% (28)	35% (21)
Males — controls	95% (114)	74% (89)	63% (76)

As in the case of the rat groups, in order to test whether the putative high dose groups did indeed have a "slight increased mortality compared to the controls", a Cox proportional hazard regression model was fitted to survival times for the high dose group against the control group. The results are presented in Table 21.

Table 21: Proportional hazards regression of high dose group compared with control (mice)

		High dose group vs control
Females	χ^2_1	0.5918
	p-value	0.4417
	HR (CI)	1.151 $(0.804, 1.649)$
Males	χ^2_1	0.13.276
	p-value	0.0003
	HR (CI)	1.798 $(1.311, 2.465)$

In the case of the female mouse study, 52% of the high dose group survived to week 91. Therefore we are not too concerned about the possibility that this dose level is too high. The proportional hazards analysis does not indicate a significantly higher risk level than the control group, but the sharp contrast in weight gain makes it reasonable to accept this dose level as a suitable high dose level.

In the case of the male mouse study, only 35% of the animals survived to the end of week 91. However, as the dose group started with 60 animals, this mortality rate left 21 animals alive at this

Table 22: Mean proportion of weight gain relative to control group (mice)

	0.025 mg/kg	0.25 mg/kg	2.5 mg/kg
Females	87.2%	78.0%	61.5%
Males	99.5%	82.2%	81.3%

point. While we should have some concerns that this dose level might exceed the maximum tolerated dose, enough animals have survived that we may attempt meaningful analyses. The elevated hazard ratio (the confidence interval is $(1.311, 2.465)$) allows us to comfortably accept the proposition that this dose level is indeed high enough to provide a strong tumor challenge to the animals.

4.1.4 Appropriateness of intermediate doses in rat study

The dose levels in the female rat study are complicated by the difficulty in establishing a suitable level for the high dose. However, the ratios between successive dose levels are very wide, suggesting that the low dose of 0.05 mg/kg is so low as to provide little useful information. Furthermore, the difference between the high dose group and the very high dose group (a factor of 5) is problematic, especially when considering the problem of which to treat as a high dose level. For this reason, we have conducted the analyses using all four treatment groups.

The same problems attend the male rat study, although they are attenuated somewhat by that observation that 2.5 mg/kg is a satisfactory high dose level for these animals.

4.1.5 Appropriateness of intermediate doses in mouse study

While it seems that the choice of 2.5 mg/kg seems to have been a reasonable dose level for the high dose group, the choice for the other dose levels is eye catching. As a result of choosing dose ratios of 10 (both between the low and mid dose group and between the mid and high dose group) the data collected from the study is concentrated towards very low doses. Specifically, 80% of the mice in the study were receiving a dose less than 10% of the high dose. Consequently, trend statistics should be viewed as less informative than is usual in such studies. Furthermore, tests which fail to detect a difference in incidence rate between the mid or low dose groups and the controls should also be treated with more skepticism than is usual, as such dose levels are a much smaller proportion of the MTD than one would normally find in such a study.

4.2 Reporting of autolysis and missing organs

The frequency with which organs were reported as autolytic was quite high in the mouse study. Furthermore, the rate of reported missing organs was also very high. Most noteworthy, however, were the patterns of reporting. Certain organs were consistently reported as missing, or autolytic in one study, but not in other studies. This would appear to indicate differing practices between different pathologists. Accordingly, caution must be taken when comparing results from one study with another, or comparing male and female animals.

5 Summary

In this submission the sponsor included reports of two animal carcinogenicity studies, one in rats and one in mice. These studies were intended to assess the carcinogenic potential of FTY 720 in rats and mice when administered orally by gavage once daily at several dosage levels for about 104 weeks.

In this review, 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.

5.1 Rat study

Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were four treated groups and two identical control groups. Three hundred and five Wistar (Crl:(WI) BR) rats of each sex were randomly allocated to treated and control groups in equal size of 51 animals. The treatment was administered orally by gavage to groups of 51 male and 51 female rats at dosages (base) of 0.05, 0.15, 0.5 and 2.5 mg/kg/day and a volume-dosage of 2 mL/kg, once daily, 7 days a week, in general for 104 weeks. The two groups of control animals similarly received the vehicle, millipore graded water.

Treatment was stopped in high dose females in week 75 and in high dose males in week 99 at a survival level of 20 animals. One female control group was terminated together with the 2.5 mg/kg/day females in week 95, when the number of surviving animals had dropped to 15.

The validity of any inferences based on the female rat studies is severely compromised by unsuccessful dose selection. The 2.5 mg/kg dose level was above the MTD, with the result that insufficient animals survived long enough to adequately test for carcinogenicity effects. Compounding this effect, the evidence that the 0.5 mg/kg dose level was adequately close to the MTD is weak.

The slightly better survival outcomes of the male rats in the 2.5 mg/kg dose group allow us to consider the dose selection to have been somewhat more adequate for the study of male animals.

The only detectable carcinogenicity signal in the rat study was for an increased incidence of thyroidal c-cell adenomas among male rats. The signal was very strong when contrasting the low dose rats with the controls, and was also detectable when contrasting the mid dose rats with control group. The signal was not detectable when contrasting the high and very high dose levels with the controls, and was not detectable among the female rats. Nonetheless, given the strength of the signal ($p = 0.0004$ for low vs control and $p = 0.0378$ for mid vs control), taken together with the lack of sensitivity of such tests, we should take the possibility that this carcinogenicity signal is genuine seriously.

5.2 Mouse study

Two separate experiments were conducted, one in males and one in females. In each of these two experiments there were three treated groups and two identical control groups. Three hundred Crl:CD-1^{TM}(ICR)BR mice of each sex were randomly allocated to treated and control groups of 60 animals each. The treatment was administered orally by gavage. The dose levels for treated groups were 0.025, 0.25, and 2.5 mg/kg/day. In this review these dose groups were referred to as the low, mid, and high dose groups, respectively. The vehicle for the test was a solution of graded millipore water; animals received a volume dosage of 5 ml/kg per day. Treatment was daily for about 104 weeks, except for the high dose males, for which treatment was stopped after 92 weeks.

One of the female mouse control groups experienced significantly higher mortality than the other (CI for hazard ratio: $(1.145, 1.562)$, $p = 0.0002$). Accordingly, the worse performing group has been stripped from the analysis of survival data. Nonetheless, the existence of such a discrepancy raises the concern that the study has been insufficiently controlled. If some external factor is affecting one control group, but not the other, then we cannot assess how this factor is affecting the various treatment groups.

Inferences based on this study are called into further question by the choice of dose levels. While 2.5 mg/kg does seem to have been a reasonable level for the high dose group, both for male and female mice, the decision to set the mid dose level at just 10% of this level, and the low dose level at 1%, means that we should not place too much weight on the data from these dose groups. In particular, trend statistics should be viewed with more skepticism than normal.

Analysis of tumor data indicated a strong signal for malignant lymphoma in mice. The evidence of a carcinogenic effect in male mice is very strong, and while the evidence from the female mice is less strong, it provides strong corroboration for the effect. Among male mice, the adjusted incidence rate (found by weighting an animal at risk by the cube of its age) was 9.9% in the control group, 11.6% in the low dose group, 23.4% in the mid dose group, and 44.2% in the high dose group. The trend is strongly significant ($p < 0.0001$). Among female mice, the corresponding incidence rates are 27.1%, 24.3%, 52.7% and 47.9%, with the differences between the mid and low dose groups and the control both being significant at the 5% level. The trend is also significant ($p = 0.0293$) at the 0.05 level.

Significant signals were also detected in the male mice for hepatic hemangiosarcomas and myeloid leukemia.

In the case of the hemangiosarcomas, the trend is significant at the 0.005 level, and the p-value for the comparison between the high dose and control dose (0.0051) is significant at the $\alpha = 0.05$ level. The signal is not replicated in the female mice, but it should be noted that only one female mouse, an animal in the high dose group, was found to have developed a hepatic hemangiosarcoma. While the evidence of carcinogenicity is therefore not as strong as for malignant lymphoma, this signal should still be considered seriously.

In the case of myeloid leukemia, the significant result is driven by the four male mice who developed tumors in the mid dose group (an adjusted rate of 9.5%, compared to 1.1% in the control group, 4.8% in the mid dose group, and 0% in the high dose group). In light of the appropriate multiplicity adjustment, and the absence of any signal among the female mice, this dose not constitute strong evidence for a carcinogenicity effect.

Note that in light of the discrepancies in reporting missing and autolytic organs, we are not justified in making comparisons in the tumor rates between male and female mice. At best we can use the one sex as a verification sample to investigate potential signals detected in the other sex.

References

[1] P. Armitage. Tests for linear trends in proportions and frequencies. *Biometrics*, 11(3):375–386, 1955.

[2] A J Bailer and C J Portier. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*, 44(2):417–31, 1988.

[3] G S Bieler and R L Williams. Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics*, 49(3):793–801, 1993.

[4] KC Chu, C Cueto, and JM Ward. Factors in the evaluation of 200 National Cancer Institute carcinogen bioassays. *Journal of toxicology and environmental health*, 8(1-2):251–280, 1981.

[5] J K Haseman. A reexamination of false-positive rates carcinogenesis studies. *Fundamental and applied toxicology*, 3(4):334–339, 1983.

[6] E.L. Kaplan and P. Meier. Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[7] K K Lin and M A Rahman. Overall false positive rates in tests for linear trend in tumor incidence in animal carcinogenicity studies of new drugs. *J Biopharm Stat*, 8(1):1–15; discussion 17–22, 1998.

[8] R Peto, M C Pike, N E Day, R G Gray, P N Lee, S Parish, J Peto, S Richards, and J Wahrendorf. Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. *IARC Monogr Eval Carcinog Risk Chem Hum Suppl*, NIL(2 Suppl):311–426, 1980.

[9] M A Rahman and K K Lin. A comparison of false positive rates of peto and poly-3 methods for long-term carcinogenicity data analysis using multiple comparison adjustment method suggested by Lin and Rahman. *J Biopharm Stat*, 18(5):949–58, 2008.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-22527	ORIG-1	NOVARTIS PHARMACEUTICALS CORP	FINGOLIMOD HCL ORAL CAPSULES

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

/s/

MATTHEW T JACKSON
06/25/2010

KARL K LIN
06/25/2010
Concur with review