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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 202,992 / S13
Drug Name:	Aubagio® (Teriflunomide)
Indication(s):	Multiple Sclerosis
Applicant:	Sanofi
Date(s):	Date of Submission: November 2, 2020 PDUFA Date: May 2, 2021
Review Priority:	Priority Review
Biometrics Division:	Division I
Statistical Reviewer:	Sharon Yan, PhD
Concurring Reviewers:	John Lawrence, PhD, Acting Team Leader Jim Hung, PhD, Director
Medical Division:	Neurology II
Clinical Team:	Laura Baldassari, MD, Clinical Reviewer Paul Lee, MD, Team Leader and Deputy Director Nicholas Kozauer, MD, Director
Project Manager:	Rania Younes, PharmD

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Teriflunomide (Aubagio®) is currently indicated for the treatment of adult patients with relapsing forms of multiple sclerosis (RMS) in the United States.

(b) (4)

Study EFC11759 was a multicenter, randomized, double-blind, placebo-controlled, parallel-group study followed by an open-label (OL) extension. A total of 166 patients were randomized and treated in the double-blind period: 109 in the teriflunomide group and 57 in the placebo group. The primary endpoint was time to first confirmed clinical relapse.

The numerical reduction of Teriflunomide in the risk of clinical relapse by 34% relative to placebo did not reach statistical significance (p = 0.29). Study EFC11759 therefore did not meet its primary endpoint.

2 INTRODUCTION

2.1 Overview

The basis for this (b) (4) submission is the efficacy and safety results of the phase-3 study EFC11759: a two-year, multicenter, randomized, double-blind, placebo controlled, parallel group trial to evaluate efficacy, safety, tolerability, and pharmacokinetics of teriflunomide in pediatric patients with RMS. The primary endpoint was time to first confirmed clinical relapse (CCR).

In the primary analysis, the risk reduction of Teriflunomide in confirmed clinical relapse, estimated at 34% relative to placebo, did not reach statistical significance (p = 0.2949). In a sensitivity analysis using a composite endpoint of time to first CCR or high MRI activity, the estimated reduction in the combined risk for teriflunomide was 43% relative to placebo (p=0.0409).

While the double-blind study was ongoing, it was noted based on blinded information that more patients (21 patients) than anticipated were switching to the open-label (OL) period, two-thirds of them (14 patients) due to high MRI activity. Given the anticipated impact on the primary analysis, (b) (4) along with the confirmed clinical relapses in a composite primary endpoint that corresponded to a preplanned key sensitivity analysis, was discussed with the EMA and the FDA. (b) (4) The FDA indicated that the preplanned sensitivity

analyses of the composite primary endpoint would be informative and considered in the review (FDA Written Responses for a Type C Meeting Request dated 18 January 2019).

The following table presents a summary of the studies included in this review.

Table 1 List of All Studies Included in This Review

	Phase and Design	Treatment Period	Comparator	# of Subjects randomized	Study Population
EFC11759	Phase 3, randomized, double-blind, placebo-controlled	The DB period is up to 96 weeks	Placebo	57 to placebo; 109 to teriflunomide	Pediatric patients 10-<18 years old with RMS

Source: Reviewer's summary

2.2 Data Sources

Original supplemental NDA July 24, 2020: \\CDSESUB1\evsprod\NDA202992\0213
Resubmitted supplemental NDA November 2, 2020: \\CDSESUB1\evsprod\NDA202992\0221

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

No notable issues were identified in the submission of data and study documents.

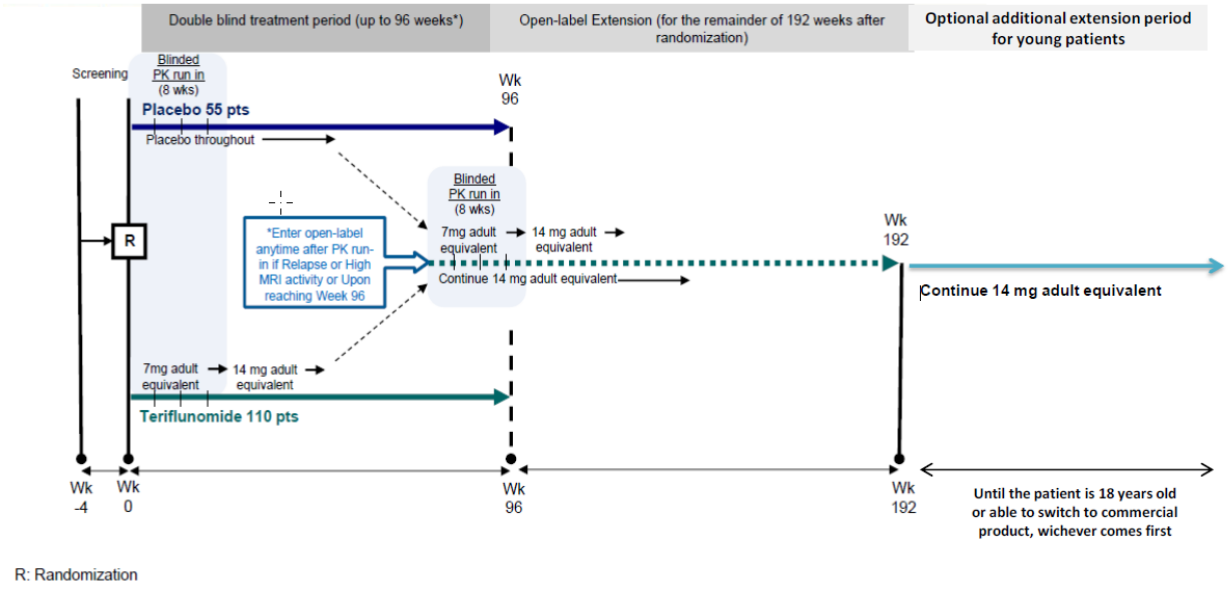
3.2 Evaluation of Efficacy

3.2.1 Evaluation of Efficacy for Study EFC11759

3.2.1.1 Study Design

Study EFC11759 was a multicenter, randomized, double-blind, placebo-controlled, parallel-group study. The study consisted of a screening period (up to 4 weeks), followed by a double-blind treatment period (up to 96 weeks), and an open-label treatment period. The double-blind period included a blinded PK run-in (8 weeks) phase consisting of 4 weeks of PK sample collection plus 4 weeks of analysis. Patients were randomized to either teriflunomide or placebo in a 2:1 ratio (110 patients on teriflunomide versus 55 patients on placebo were planned).

The study design is graphically described in Figure 1.



R: Randomization

Figure 1 Graphical Study Design (source: Protocol)

3.2.1.2 Study Endpoints

Primary Endpoint

The primary efficacy endpoint was the time to first confirmed clinical relapse (CCR) after randomization up to the end of double-blind treatment period. Clinical relapses were reviewed for confirmation by an independent relapse adjudication panel (RAP).

Secondary Endpoint

Secondary efficacy endpoints were the following:

- Proportion of clinical relapse-free patients at 24, 48, 72 and 96 weeks
- MRI endpoints based on central reading:
  - Number of new/newly enlarged T2 lesions
  - Number of Gadolinium (Gd)-enhancing T1 lesions
  - Change in volume of T2 lesions
  - Change in volume of T1 hypointense lesions
  - Number of new hypointense T1 lesions
  - Proportion of patients free of new or enlarged MRI T2 lesions at 48 weeks and 96 weeks
  - Percentage change of brain volume.
- Cognitive outcome measured by the symbol digit modalities test (SDMT) and Cognitive Battery Tests

The numbers of new/newly enlarged T2 lesions and the numbers of Gd-enhancing T1 lesions were considered the key secondary endpoints.

MRI was performed at Weeks 24, 48, 72 and 96. In case of at least 5 new/enlarged T2 lesions at the MRI of Week 24, an additional MRI was performed at Week 36. The patients then had the option to continue into the open-label period early to receive teriflunomide treatment in case of high MRI activity defined as:

- At least 9 new/enlarged T2 lesions at Week 36, or,
- At least 5 new/enlarged T2 lesions on each of the 2 consecutive MRI scans of Week 36 and Week 48, or,
- At least 5 new/enlarged T2 lesions on each of the 2 consecutive MRI scans of Week 48 and Week 72.

3.2.1.3 Statistical Methodologies

3.2.1.3.1 Analyses of the Primary Endpoints

The primary efficacy endpoint of time to first confirmed clinical relapse (CCR) was to be analyzed using a stratified log-rank test with time to first CCR as the dependent variable, treatment group as a test variable, and region and baseline pubertal status as covariates. Treatment effect as measured by the hazard ratio and its associated 95% confidence interval (CI) was to be estimated using a Cox's proportional-hazards model with robust variance estimation. The Cox's model was to include the factors for treatment group, region, baseline pubertal status, age at study entry, and number of relapses in the year prior to randomization.

The following sensitivity analyses using the similar log-rank test and Cox proportional-hazards model as described above were planned for the primary endpoint:

- Time to first CCR or high MRI activity meeting protocol criteria for switching into open-label period, whichever came first;
- Time to first clinical relapse (i.e., clinical relapse confirmed or not);
- Time to first CCR occurring after the PK run-in (8 weeks) phase. Patients who have a relapse during the PK run-in (8 weeks) phase will be included in the analysis with the time to first clinical relapse right censored at the time of treatment discontinuation;
- Time to first clinical relapse with objective signs on the Examining Neurologist's examination including relapses during the PK run-in (8 weeks) phase and relapses reported after the study drug discontinuation and up to 96 weeks after randomization.

3.2.1.3.2 Analysis of Secondary Endpoints

The proportion of patients free of confirmed clinical relapse at Weeks 24, 48, 72 and 96 were to be estimated based on Kaplan-Meier methods. A Kaplan-Meier graph summarizing the event probability over time was to be presented.

The number of new or enlarged T2 lesions per MRI scan was to be analyzed using a negative binomial regression model with robust variance estimation. The model was to include the total

number of new or enlarged T2 lesions as the response variable, with treatment group, region, pubertal status and age as covariates, and the log-transformed number of scans as an offset variable. The estimated number of lesions per scan and associated 2-sided 95% confidence interval (CI) were to be provided for each treatment group.

The number of T1 Gd-enhancing lesions per MRI scan was to be analyzed using a similar negative binomial regression model as described above for T2 lesions.

To strongly control Type-I error rate for this family, a step-down testing procedure was to be applied to the 2 key secondary efficacy endpoints in the order specified below.

- Number of new/newly enlarged T2 lesions
- Number of Gd-enhancing T1 lesions

Each hypothesis was to be formally tested only if the preceding one was significant at 5% level.

3.2.1.3.3 Pooling of Study Centers

Study centers were to be pooled into geographical regions for statistical analysis as follows.

- Europe: Belgium, Bulgaria, Estonia, France, Greece, Lithuania, Netherlands, Portugal, Russian federation, Serbia, Spain, Ukraine, United Kingdom
- North America: Canada, United States
- Asia: China
- Middle East: Turkey, Israel, Lebanon
- North Africa: Tunisia, Morocco

3.2.1.4 Study Patients

3.2.1.4.1 Patient Disposition

A total of 185 patients were screened for entry eligibility and 166 patients were randomized to study treatment: 109 patients in the teriflunomide group and 57 patients in the placebo group. All randomized patients were treated and 93.6% and 93.0% of the randomized patients completed the double-blind period in the teriflunomide and placebo groups, respectively.

A higher proportion of patients met the criteria of high MRI activity in the placebo group than in the teriflunomide group (26.3% and 12.8%, respectively) and a higher proportion of patients experienced a CCR in the placebo group than in the teriflunomide group (38.6% and 29.4%, respectively). Consequently, the proportion of patients who completed the 96-week double-blind period was higher in the teriflunomide group than in the placebo group (51.4% and 28.1%, respectively).
The main reason for permanent treatment discontinuation was AE in the teriflunomide group (5.5%) and lack of efficacy or other reason in the placebo group (3.5%).

A summary of patient disposition is presented in Figure 2.

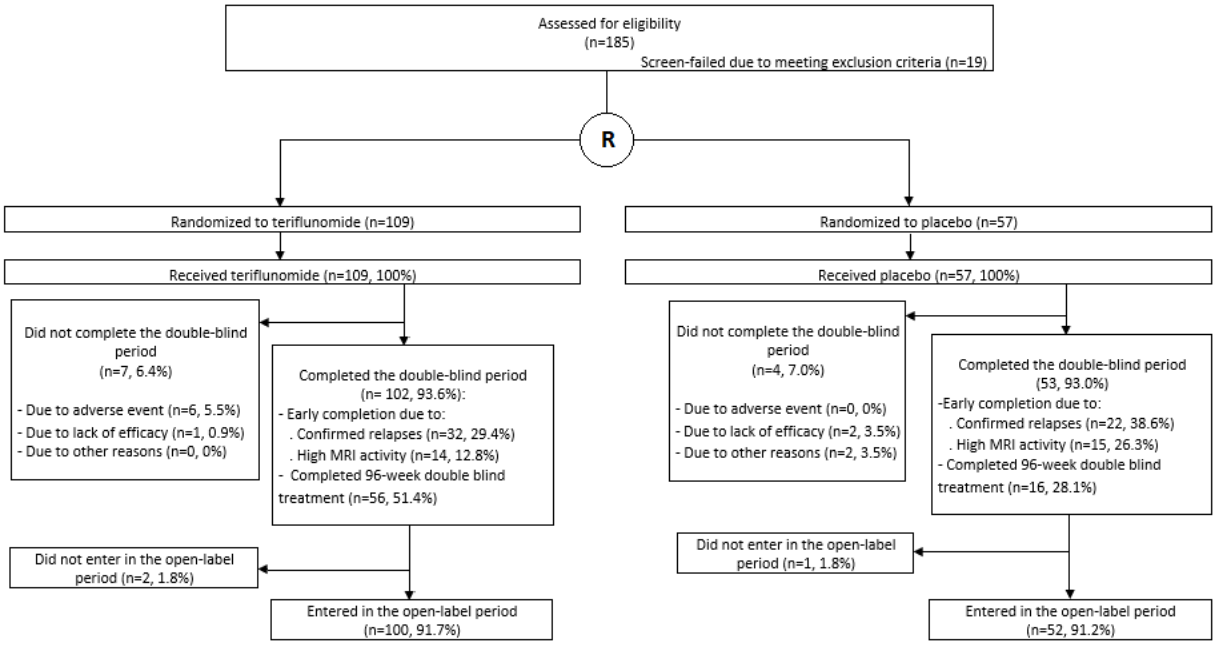


Figure 2 Disposition of Subjects (Source: Figure 2 of CSR)

3.2.1.4.2 Patient Demographics

Patient demographics were balanced between treatment groups. The mean age of the randomized population was 14.6 years. Overall, 26 patients (15.7%) were below 13 years old at enrollment. At baseline, 10 patients (6.0%) were pre-pubertal (Tanner Stage I): 4.6% of patients in the teriflunomide group and 8.8% in the placebo group. Most patients were female (66.9%); 70.5% of patients were Caucasian/White and 22.3% were Asian/Oriental.

3.2.1.4.3 Patient Baseline Disease Characteristics

Disease characteristics at baseline were generally similar among treatment groups (Table 2). All patients were diagnosed with relapsing remitting MS. The median time since first diagnosis of MS was 0.69 years, and the median time since first symptoms of MS was 1.61 years. All patients experienced relapses within the past 2 years (median of 2 relapses). Overall, 19.9% of patients received MS medications in the last 2 years. At baseline, about half of the patients in both groups had ≥1 T1 Gd-enhancing lesions on MRI. The mean patient level T1 Gd-enhancing lesions was 3.9 lesions.

Table 2 Baseline disease characteristics - randomized population

	Placebo N=57	**Teriflunomide N=109**
Time since diagnosis of MS, year		
Mean (SD)	1.40 (1.71)	1.40 (1.81)
Median	0.70	0.68

Time since most recent relapse onset, month		
Mean (SD)	5.79 (4.04)	4.97 (3.05)
Median	4.99	4.27
Number of relapses past 1 year		
Mean (SD)	1.4 (0.7)	1.6 (0.7)
Median	1.0	1.0
Min; Max	0; 3	1; 4
Number of relapses past 2 years		
Mean (SD)	2.0 (1.0)	2.1 (1.0)
Median	2.0	2.0
Min; Max	1; 6	1; 5
With MS medication last 2 years, n (%)		
Yes	14 (24.6)	19 (17.4)
No	43 (75.4)	90 (82.6)
Number of T1 Gd lesions		
Mean (SD)	3.9 (7.74)	3.9 (7.50)
Median	1.0	1.0
Min; Max	0; 38	0; 39

Source: Table 12 of CSR

3.2.1.5 Efficacy Results

3.2.1.5.1 Primary Endpoint – Time to First Confirmed Clinical Relapse

Confirmed clinical relapse (CCR) occurred in 36.7% of the patients in the teriflunomide group and 43.9% in the placebo group during the double-blind period. The log-rank test of the time to first CCR resulted in a p-value of 0.2949. Therefore, the study did not meet its primary endpoint.

The estimated probability of CCR at Week 96, using the Kaplan-Meier method was 0.531 in the placebo group and 0.389 in the teriflunomide group. The estimated hazard ratio of teriflunomide to placebo was 0.657, corresponding an estimated relative risk reduction of 34.3%, from the Cox proportional hazard model.

A sensitivity analysis was performed using the time to first CCR or high MRI activity meeting criteria for switching into the open-label period, whichever came first. Fifteen patients in each treatment group had defined high MRI activity. Among them, one patient in each treatment group had both CCR and high MRI activity. The reduction of Teriflunomide in the combined risk of CCR or high MRI activity as compared to placebo was estimated at 43.4% with a p-value of 0.0409. The following table provide a summary of analysis of relapses.

Table 3 Analysis of confirmed clinical relapse and high MRI activity

	Placebo	**Teriflunomide**

	N=57	N=109
Primary Analysis		
Number (%) of patients with CCR	25 (43.9)	40 (36.7)
Estimated probability of CCR		
At 48 weeks	0.391	0.298
At 96 weeks	0.531	0.389
Hazard Ratio (95% CI)		0.657 (0.388, 1.113)
p-value from log-rank test		0.2949
Sensitivity analysis		
Number (%) of patients with CCR or high MRI activity	39 (68.4)	54 (49.5)
Hazard ratio (95% CI)		0.566 (0.368, 0.870)
p-value		0.0409

Source: reviewer's analysis

An inspection was performed for the two Turkish sites. There was one subject in one of the sites who appeared to have had a MS relapse within 30 days of randomization. The Office of Scientific Investigations recommended sensitivity analysis excluding the subject. The p-value from the sensitivity analysis excluding the subject was 0.3049 with estimated hazard ratio (95% CI) of 0.661 (0.390, 1.119).

Results from other sensitivity analysis of time to first CCR or time to first clinical relapse (confirmed or not) were similar to the primary analysis results.

The Kaplan-Meier plot of time to first CCR in the double-blind period is presented in Figure 3.

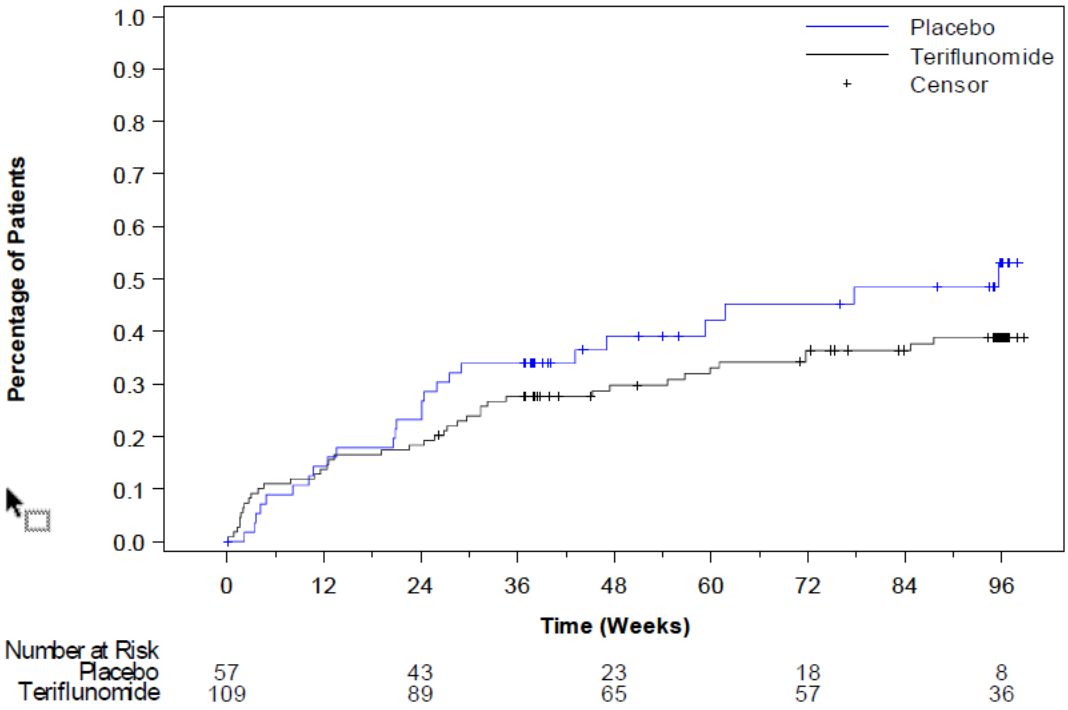


Figure 3 Kaplan_Meier plot of time to first confirmed relapse
Source: Figure 3 of CSR

Reviewer's Comments to Sensitivity Analysis

The criteria for high MRI activity were not set for efficacy reasons and were not set consistently throughout the treatment period. The criteria could be met by having at least 5 lesions at two consecutive scans at Weeks 36 and 48 or Weeks 48 and 72, but not at Weeks 24 and 36 or Weeks 72 and 96. It required a minimum of 5 lesions at Week 24 and 9, instead of 5, lesions at Week 36 and no subject could meet the criteria at Weeks 72 and 96 regardless of the number of lesions. In addition, the following examples were found in the data:

- There were patients who had > 5 lesions at Week 24 without a Week 36 scan. It is not clear why the Week 36 scan was not performed.
- There were patients who had > 5 lesions at Weeks 24 and 48 without Week 36 scan. In one such case, a subject was indicated as having high MRI activity and in another case a subject was not indicated as having high MRI activity.

The following table presents mean and medians of the number of new/enlarging T2 lesions in subjects who had at least one MRI scan with respect to subject status of having CCR or not.

Table 4 Number of new/enlarging T2 lesions by status of CCR (subjects with ≥ 1 MRI scan)

Number of new/enlarging T2 lesions	Placebo N=45	Teriflunomide N=100

Subjects with CCR		
N	14	31
Mean	12.8	9.2
Median	2.7	7.0
Subjects without CCR		
N	31	69
Mean	20.0	6.3
Median	5.8	3.0

Source: Reviewer's analysis

In the placebo group, the mean and median of new/enlarging T2 lesions were larger in patients who did not have CCR than in patients who had CCR. These number were in opposite direction in the teriflunomide group. Thus, having a larger number of new/enlarging T2 lesions is not indicative of higher likelihood of CCR. Therefore, a larger number of subjects who switch to the open-label treatment in the placebo group than in the teriflunomide group does not add evidence of efficacy in terms of the primary endpoint other than those shown in the primary analysis.

3.2.1.5.2 Key Secondary Endpoints

The two key secondary endpoints were the number of new/newly enlarged T2 lesions and the number of Gd-enhancing T1 lesions.

Analysis of the number of new or enlarging T2 lesions

Forty five of the 57 placebo-treated patients and 100 of the 109 teriflunomide-treated patients had at least one post-baseline MRI performed. Among these patients, 38 (84.4%) patients in the placebo group and 85 (85%) patients in the teriflunomide group had at least 1 new/enlarging T2 lesions. The median number of new/enlarging T2 lesions per scan was 5.0 for the placebo group and 4.1 for the teriflunomide group. After adjusting for age, region and pubertal status, the estimated new/enlarging T2 lesion numbers per scan was 10.515 for the placebo group and 4.735 for the teriflunomide group. The risk reduction in the new/enlarging T2 lesions for teriflunomide was 55% (p=0.0006) as compared to placebo. Analysis results are presented in Table 5.

Table 5 Analysis of new/enlarging T2 lesions and T1 Gd-enhancing lesions

	Placebo N=57	**Teriflunomide N=109**
New/enlarging T2 lesions		
Number of patients with MRI scans	45	100
Patients with new/enlarging T2, n (%)	38 (84.4)	85 (85.0)

Mean (SD) per scan	17.8 (26.3)	7.2 (9.3)
Median per scan	5.0	4.1
Min; Max per scan	0; 134	0; 42
Adjusted number new/enlarging T2 per scan		
LS mean (95% CI)	10.515 (4.705, 23.500)	4.735 (2.122, 10.567)
Relative risk (95% CI)		0.45 (0.285, 0.711)
p-value		0.0006
T1 Gd-enhancing lesions		
Patients with post-baseline MRI scans, n	45	100
Baseline		
N	45	98
Patients with T1 Gd lesions, n (%)	24 (53.3)	53 (53.0)
Mean (SD)	3.5 (6.88)	3.6 (6.42)
Median	1.0	1.0
Min; Max	0; 35	0; 36
Post-baseline		
Patients with T1 Gd lesions, n (%)	32 (71.1)	55 (55.0)
Mean (SD) per scan	5.1 (11.7)	1.4 (3.6)
Median per scan	0.8	0.2
Min; Max per scan	0; 56	0; 26
LS mean (95% CI)	7.505 (2.482, 22.695)	1.897 (0.656, 5.489)
Relative risk (95% CI)		0.235 (0.126, 0.505)
p-value		<0.0001

Source: reviewer's analysis

Analysis of the number of T1 Gd-enhancing lesions

At baseline, treatment groups were balanced in the Gd-enhancing lesion numbers; 24 (53.3%) of the 45 patients in the placebo group and 53 (53.0%) of the 100 patients in the teriflunomide group had at least one Gd-enhancing lesions. The median number of Gd-enhancing lesions was 1.0 for both treatment groups.

Post-baseline, the median number of T1 Gd-enhancing lesions per scan was 0.8 for the placebo group and 0.2 for the teriflunomide group. After adjusting for age, region, pubertal status and baseline T1 lesion numbers, the estimated mean for the number of T1 Gd-enhancing lesions per scan was 7.505 for the placebo group and 1.897 for the teriflunomide group, representing a risk reduction of 75% ($p<0.0001$) for the teriflunomide compared to placebo (Table 5).

3.3 Evaluation of Safety

Please refer to Evaluation of Safety by Dr. Laura Baldassari for evaluation of safety.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

Analysis of time to first CCR was performed for subpopulations of sex, age group, race, region, and pubertal status (Table 6). Age was grouped by < 13 years and ≥ 13 years as it closely divided pubertal status. There were only 7 African American subjects and 5 subjects in the "Other" race in total and the numbers were too small to make a meaningful analysis. Therefore, analysis by race was only performed on Asian and White subgroups. Due to the small number of subjects in regions of North America (6 subjects) and North Africa (7 subjects), these two regions were not included in analysis by region.

Results from the analysis of subgroup populations were generally consistent with the overall population.

Table 6 Analysis of confirmed clinical relapse by subgroups of sex, age group and region

	Placebo N=57	**Teriflunomide N=109**
Sex		
Male		
N	18	37
Number (%) with CCR	8 (44.4)	12 (32.4)
Hazard Ratio (95% CI)		0.668 (0.274, 1.629)
Nominal p-value		0.5122
Female		
N	39	72
Number (%) with CCR	17 (43.6)	28 (38.9)
Hazard Ratio (95% CI)		0.743 (0.387, 1.431)
Nominal p-value		0.5116
Age Group		
< 13 years		
N	10	16
Number (%) with CCR	6 (60.0)	2 (12.5)
Hazard Ratio (95% CI)		0.142 (0.024, 0.853)
Nominal p-value		0.0477*
≥ 13 years		
N	47	93
Number (%) with CCR	19 (40.4)	38 (40.9)
Hazard Ratio (95% CI)		0.814 (0.448, 1.479)
Nominal p-value		0.7370
Race		
White		
N	42	75
Number (%) with CCR	16 (38.1)	25 (33.3)
Hazard Ratio (95% CI)		0.684 (0.345, 1.353)
Nominal p-value		0.4678
Asian		
N	12	25
Number (%) with CCR	8 (66.7)	9 (36.0)

Hazard Ratio (95% CI)		0.473 (0.182, 1.233)
Nominal p-value		0.1435
Region		
Asian	12	25
N	8 (66.7)	9 (36.0)
Number (%) with CCR		0.473 (0.182, 1.233)
Hazard Ratio (95% CI)		0.1435
Nominal p-value		
Europe		
N	24	47
Number (%) with CCR	11 (45.8)	21 (44.7)
Hazard Ratio (95% CI)		0.742 (0.342, 1.613)
Nominal p-value		0.7357
Middle East		
N	16	29
Number (%) with CCR	5 (31.3)	8 (27.6)
Hazard Ratio (95% CI)		0.842 (0.226, 3.135)
Nominal p-value		0.6210
Pubertal Status		
Prepubertal		
N	5	5
Number (%) with CCR	4 (80.0)	0 (0)
Hazard Ratio (95% CI)		0
Nominal p-value		0.0134
Pubertal		
N	52	104
Number (%) with CCR	21 (40.4)	40 (38.5)
Hazard Ratio (95% CI)		0.790 (0.454, 1.376)
Nominal p-value		0.6009

*The test from which the nominal p-value was obtained is questionable due to lack of events in some strata.
Source: Reviewer's analysis

4.2 Other Special/Subgroup Populations

No other subgroups were analyzed.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

No notable statistical issues were found.

5.2 Collective Evidence

Collective evidence was not evaluated since this was a single study application.

5.3 Conclusions and Recommendations

Study EFC11759 failed to provided evidence that teriflunomide was efficacious in delaying the MS relapses in pediatric RMS patients aged 10-17 as compared with placebo.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

XIAORONG YAN
04/02/2021 10:24:07 AM

JOHN P LAWRENCE
04/02/2021 05:21:25 PM

HSIEN MING J HUNG
04/02/2021 10:20:16 PM